

3D-QSAR/HQSAR-based analysis of bioconcentration and molecular modification of monophenyl aromatic compounds

Qing LI^{1,2}, Wenwen GU^{1,2}, Yu LI^{1,2,*}¹College of Environmental Science and Engineering, North China Electric Power University, Beijing, P.R. China²State Key Laboratory of Regional Optimisation of Energy System, North China Electric Power University, Beijing, China

Received: 11.07.2018

Accepted/Published Online: 28.11.2018

Final Version: 05.02.2019

Abstract: Comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) of three-dimensional quantitative structure–activity relationship (3D-QSAR) models were conducted using 36 aromatic compound congeners as the training set and eight aromatic compound congeners as the test set. Logarithms of experimental bioconcentration data (logBCF) were used as dependent variables and the physicochemical properties of the aromatic compound congeners were used as independent variables. Furthermore, the molecular holographic quantitative structure–activity relationship (HQSAR) model constructed by the HQSAR module in Sybyl software was used to establish a method to find molecular activity substitution sites of bioconcentration potential by coupling the molecule activity contribution map of HQSAR with the electrostatic contour map of CoMSIA. Pentachlorobenzene (PeCB) was used as the target molecule to design PeCB derivatives with low bioconcentrations. Seven PeCB derivatives with low bioconcentrations showed no significant changes in molecular toxicity and migration, while 5-NO₂-PeCB and six other PeCB derivatives showed increased degradation abilities. By quantitative calculation using density functional theory, the flame retardancy and insulation properties of the PeCB derivatives were found to remain unchanged, except for those of 5-NO₂-PeCB.

Key words: Aromatic compounds, bioconcentration, three-dimensional quantitative, holographic quantitative structure–activity relationship

1. Introduction

Aromatic compounds are substances containing one or more phenyl rings with planar or near-planar conjugated structures, including benzene, polycyclic aromatic hydrocarbons (PAHs), and their derivatives. Aromatic compounds are highly toxic, stable, and widely found in nature. Some aromatic compounds have been listed as persistent organic pollutants by the Stockholm Convention, including pentachlorobenzene (PeCB),¹ hexachlorobenzene,² polychlorobiphenyls,³ and polychlorinated naphthalenes.⁴ Furthermore, nitrobenzene and PAHs have been listed as priority pollutants by the US Environmental Protection Agency.⁵

As an important class of chemical raw materials and intermediates, aromatic compounds are widely used in food, dyes, petroleum, plastics, organic synthesis, and other industries.^{6,7} For example, PeCB has been used as an intermediate in the synthesis of some insecticides and has certain applications in flame retardants and insulating fluids. However, PeCB exhibits high enrichment, high toxicity, long-range migration, and hard-

*Correspondence: liyuxx8@hotmail.com

to-degrade POP characteristics. Owing to the environmental problems, PeCB production has ceased. Waste combustion was a main pathway for PeCB entering the environment.^{8–10} In 2005, Jaward et al. detected the presence of PeCB in the Italian Alps with a content of about 20 pg m⁻³.¹¹

Bioconcentration factors (BCFs) can reflect the balance between the concentration of a chemical substance in an organism or a certain tissue and that in the surrounding environment. BCFs can be used as important indicators of the concentration of chemical substances in organisms.¹² Owing to their stable chemical properties, high accumulation, and hard-to-degrade properties, aromatic compounds are difficult to remove from the body. Weisglas et al. reported that perinatal exposure to polychlorinated biphenyls (PCBs) and dioxins was associated with immune changes in Dutch preschool children.¹³ Furthermore, Loomis found that Americans involved in the production of electrical insulators (including PCBs) had a 4.81-fold higher probability of skin cancer and malignant melanoma than the control group (exposure time, $\geq 10,000$ h).¹⁴ Therefore, studying the enrichment of related organic pollutants and development of low bioenrichment alternatives is of great importance.

Fernández et al. used an optimized gas chromatography–mass spectrometry method to determine the bioconcentration factors of heterogeneous PAHs in hairy worms.¹⁵ Togunde used solid-phase microextraction and liquid chromatography–tandem mass spectrometry to determine the enrichment of fluoxetine and sertraline in fish such as rainbow trout.¹⁶ Torreiro-Melo et al. measured the enrichment concentration of phenanthrene in fish bile using a fluorescence method and found that exposure to a certain concentration of phenanthrene affected their swimming and foraging ability.¹⁷ The quantitative structure–activity relationship (QSAR) method has been widely used in academia and industry as an important tool in drug molecule design, environmental toxicology, and other fields.¹⁸ The study of compound enrichment and low bioconcentration design can be accomplished using QSAR methods. Cichero et al. used the 3D-QSAR method to model 52 pyrazole derivatives that are HIV-1 nonnucleoside reverse transcriptase inhibitors.¹⁹ Accordingly, the predictive power of this model can facilitate the synthesis of novel analogues. Sharma et al. established a 3D-QSAR model for 40 6-benzimidazole derivatives that are nonpeptide angiotensin II receptor antagonists.²⁰ The information provided by the model provided a better understanding of the structural requirements of antihypertensive activity and helped design new effective molecules. Although the 3D-QSAR model has been widely used as an efficient and convenient method, the force field shown in the contour map is not precise enough to explain the activity of a specific site. Therefore, the holographic quantitative structure–activity relationship (HQSAR) model is needed to perform fixed-point analysis. This HQSAR model can generate molecular holographic descriptors, and molecular fragment analysis can elucidate the active contribution of each atom in the molecule. In this study, the molecular structure and bioconcentration (logBCF) of some monophenyl aromatic compounds have been studied using the CoMFA and CoMSIA methods. A stable and predictive 3D-QSAR model was established to analyze the influence of different force fields on logBCF. By coupling CoMSIA and HQSAR results, a novel molecular design with a low bioconcentration of a single precise site in PeCB was evaluated for toxicity, mobility, degradability, and utility. This research provides a theoretical method for designing low bioconcentration alternatives to PeCB.

2. Results and discussion

2.1. Evaluation analysis and validation of aromatic compound biological enrichment using CoMFA and CoMSIA models

2.1.1. Evaluation and analysis of the CoMFA and CoMSIA models

Evaluation parameters and molecular field contribution rates of the CoMFA and CoMSIA models were obtained using PLS calculations. As shown in Table 1, the CoMFA model had an optimum component (n) of 6, cross-validation correlation coefficient (q^2) of 0.694, non-cross-validation correlation coefficient (R^2) of 0.913, and SEE of 0.300. Furthermore, the CoMSIA model had the following values: $n = 9$, $q^2 = 0.869$, $R^2 = 0.984$, and $SEE = 0.137$. For both models, the q^2 values were higher than 0.5 and the R^2 values were higher than 0.8,²¹ which showed that the models had an improved fitting ability and predictability. Scrambling stability test parameters Q^2 and dQ^2/dr^2yy are shown in Table 1. These parameters also indicated that the models had good stability. In the CoMFA model, the contribution ratios of the steric and electrostatic fields were 45.70% and 54.30%, respectively, which showed that steric effects and electrostatic interactions influenced the logBCFs of the aromatic compounds, and that electrostatic interaction was the major contributor. In the CoMSIA model, the contributions of the steric, electrostatic, hydrophobic, hydrogen bond donor, and hydrogen bond acceptor fields were 10.10%, 38.90%, 29.40%, 10.80%, and 10.80%, respectively, which showed that electrostatic interaction had the highest contribution (38.90%) (Table 2). As shown in Table 3, the relative errors of the compounds in the CoMFA model test set were -7.28%, -9.38%, -8.48%, -5.68%, -7.26%, -9.00%, -7.62%, and -7.78%, respectively, while those in the CoMSIA model were 3.08%, -6.02%, -4.96%, 1.85%, -3.28%, 0.72%, -2.96%, and -6.47%, respectively. All relative errors were less than 10%, an acceptable range.

Table 1. Evaluation parameters of the CoMFA and CoMSIA models.

Model	q^2	n	SEE	R^2	F	r_{pred}^2	SEP	Q^2	SDEP	dq^2/dr^2yy
CoMFA	0.694	6	0.300	0.913	50.821	0.761	0.412	0.508	0.716	1.132
CoMSIA	0.869	9	0.137	0.984	175.287	0.900	0.267	0.568	0.710	1.746

Table 2. Molecular fields' contribution to aromatic compounds' BCF of the CoMFA and CoMSIA.

Model	S	E	H	D	A
CoMFA	45.70%	54.30%	-	-	-
CoMSIA	10.10%	38.90%	29.40%	10.80%	10.80%

2.1.2. Validation of the CoMFA and CoMSIA models

Using the constructed CoMFA and CoMSIA models, the activity of the test set molecules was predicted to verify the accuracy of the obtained model. According to PLS analysis under the QSAR module, the standard errors (SEPs) of the CoMFA and CoMSIA models for logBCF prediction were 0.412 and 0.267, while the external predictions (r_{pred}^2) were 0.761 and 0.900 (> 0.6),²¹ respectively. These values showed that the constructed 3D-QSAR models had good external prediction capabilities.

As shown in Figure 1 and Figure 2, the predicted values and experimental values were linear in the two models and in good agreement. All data were concentrated near the trend line, and the slopes of the CoMFA

Table 3. Predicted logBCF values of aromatic compounds through CoMFA and CoMSIA models.

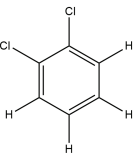
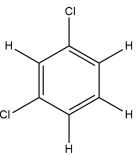
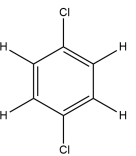
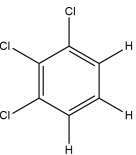
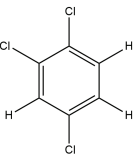
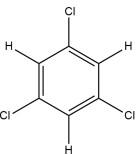
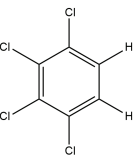
No.	Compounds	Obs.	CoMFA		CoMSIA	
			Pred.	Relative error (%)	Pred.	Relative error (%)
1 ^a	 1,2-Dichlorobenzene	3.12	2.893	-7.28%	3.216	3.08%
2 ^b	 1,3-Dichlorobenzene	3.05	2.864	-	2.894	-
3 ^a	 1,4-Dichlorobenzene	3.24	2.936	-9.38%	3.045	-6.02%
4 ^a	 1,2,3-Trichlorobenzene	3.95	3.615	-8.48%	3.754	-4.96%
5 ^a	 1,2,4-Trichlorobenzene	3.84	3.622	-5.68%	3.911	1.85%
6 ^a	 1,3,5-Trichlorobenzene	4.05	3.756	-7.26%	3.917	-3.28%
7 ^a	 1,2,3,4-Tetrachlorobenzene	4.31	3.922	-9.00%	4.341	0.72%

Table 3. Continued.

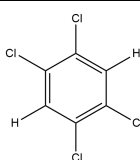
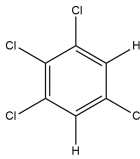
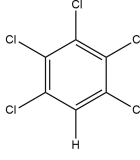
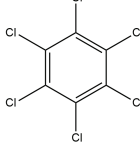
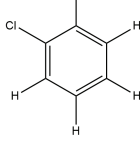
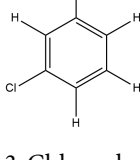
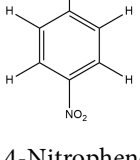
	1,2,3,4-Tetrachlorobenzene					
8 ^a	 1,2,4,5-Tetrachlorobenzene	4.67	4.314	-7.62%	4.532	-2.96%
9 ^a	 1,2,3,5-Tetrachlorobenzene	4.81	4.436	-7.78%	4.499	-6.47%
10 ^b	 Pentachlorobenzene	5.02	5.131	-	5.119	-
11 [*]	 Hexachlorobenzene	5.64	5.192	-	5.657	-
12 ^b	 2-Chlorophenol	2.54	2.345	-	2.438	-
13 ^b	 3-Chlorophenol	2.31	2.495	-	2.464	-
14 ^b	 4-Nitrophenol	1.84	2.015	-	1.87	-

Table 3. Continued.

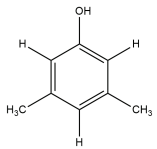
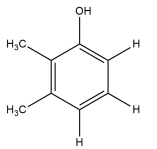
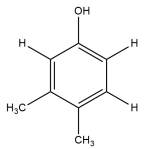
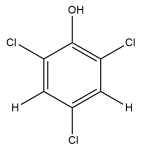
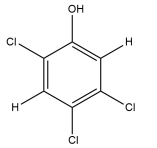
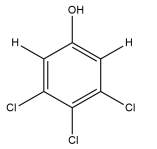
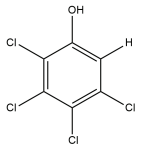
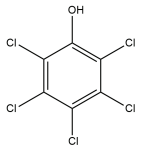
15 ^b	 3,5-Dimethylphenol	2.47	2.499	-	2.409	-
16 ^b	 2,3-Dimethylphenol	2.66	2.67	-	2.783	-
17 ^b	 3,4-Dimethylphenol	2.65	2.688	-	2.596	-
18 ^b	 2,4,6-Trichlorophenol	3.51	3.698	-	3.53	-
19 ^b	 2,4,5-Trichlorophenol	3.67	3.811	-	3.768	-
20 ^b	 3,4,5-Trichlorophenol	3.86	3.81	-	3.711	-
21 ^b	 2,3,4,5-Tetrachlorophenol	4.23	3.946	-	4.31	-
22 ^b	 1,2,3,4,5-pentachlorophenol	4.63	4.76	-	4.518	-

Table 3. Continued.

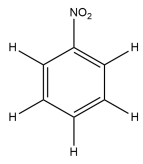
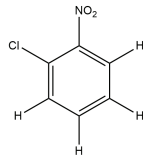
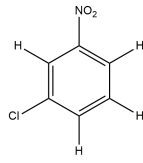
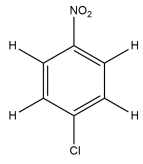
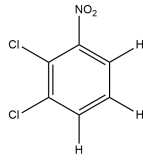
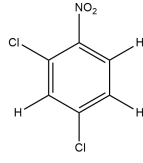
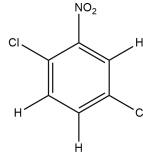
	Pentachlorophenol					
23 ^b	 Nitrobenzene	1.47	1.607	-	1.536	-
24 ^b	 o-Chloronitrobenzene	2.29	2.334	-	2.406	-
25 ^b	 m-Chloronitrobenzene	2.42	2.231	-	2.282	-
26 ^b	 p-Chloronitrobenzene	2.46	2.413	-	2.444	-
27 ^b	 2,3-Dichloronitrobenzene	3.01	3.068	-	3.114	-
28 ^b	 2,4-Dichloronitrobenzene	3.02	2.991	-	3.04	-
29 ^b	 2,5-Dichloronitrobenzene	2.92	3.159	-	3	-

Table 3. Continued.

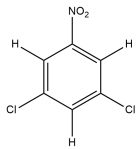
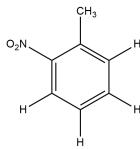
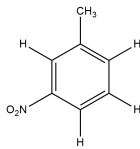
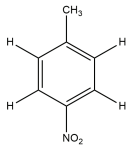
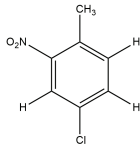
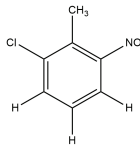
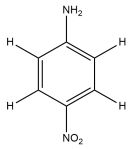
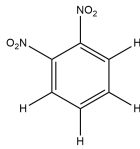
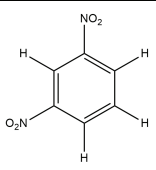
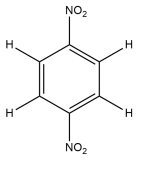
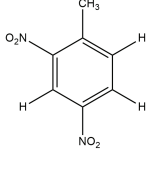
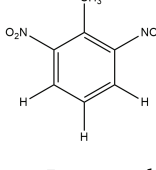
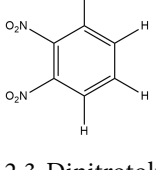

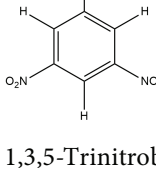
30 ^b	 3,5-Dichloronitrobenzene	3.01	2.803	-	2.913	-
31 ^b	 o-Nitrotoluene	2.28	2.056	-	2.293	-
32 ^b	 m-Nitrotoluene	2.31	2.539	-	2.327	-
33 ^b	 p-Nitrotoluene	2.37	2.136	-	2.262	-
34 ^b	 4-Chloro-2-nitrotoluene	3.02	2.963	-	2.895	-
35 ^b	 2-Chloro-6-nitrotoluene	3.09	2.796	-	3.021	-
36 ^b	 p-Nitroaniline	1.92	1.955	-	1.902	-
37 ^b		1.82	1.989	-	1.861	-

Table 3. Continued.

	o-Dinitrobenzene					
38 ^b	 m-Dinitrobenzene	1.87	2.056	-	1.781	-
39 ^b	 p-Dinitrobenzene	1.7	1.835	-	1.709	-
40 ^b	 2,4-Dinitrotoluene	2.31	2.12	-	2.447	-
41 ^b	 2,6-Dinitrotoluene	2.44	2.636	-	2.464	-
42 ^b	 2,3-Dinitrotoluene	2.49	2.67	-	2.277	-
43 ^b	 3,4-Dinitrotoluene	2.32	2.25	-	2.176	-
44 ^b	 1,3,5-Trinitrobenzene	2.95	3.007	-	2.905	-

^a Test set; ^b training set; * template molecule.

and CoMSIA models were 0.93702 and 1.00114, respectively. Furthermore, the coefficients of determination (R^2) were 0.96246 and 0.98879, respectively, indicating that the two constructed models had higher internal prediction abilities and could be used to predict single-benzene ring aromatic compound BCF.

$$y = 0.93702x + 0.17681 (R^2 = 0.96246) \quad (1)$$

$$y = 1.00114x - 0.01567 (R^2 = 0.98879) \quad (2)$$

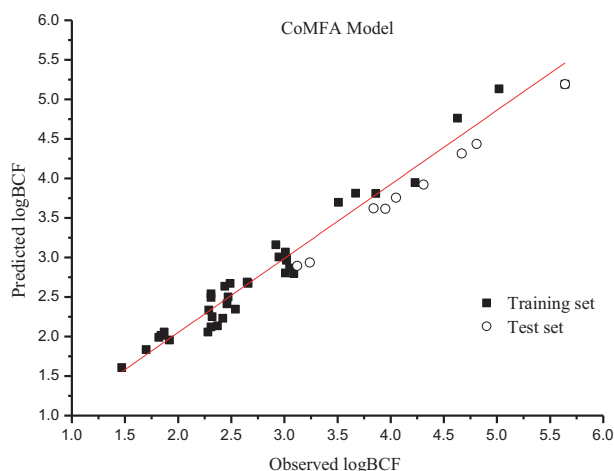


Figure 1. The plot of observed vs. predicted logBCF values using CoMFA model.

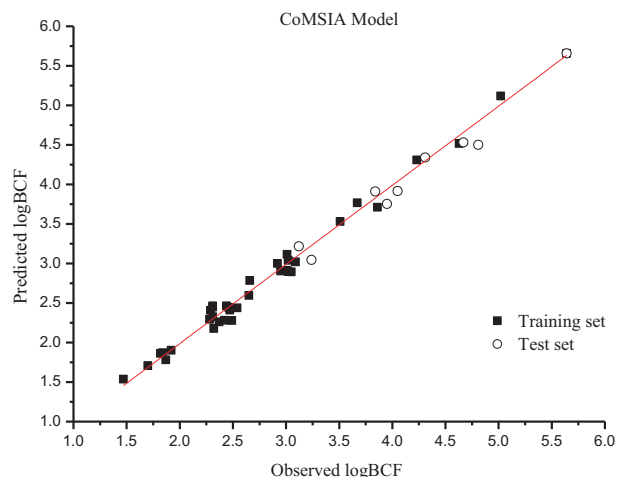


Figure 2. The plot of observed vs. predicted logBCF values using CoMSIA model.

2.1.3. Difference analysis of CoMFA and CoMSIA models

The statistical parameters of the models and the fitting degree of the validations gave R^2 values for the CoMFA and CoMSIA models that were higher than 0.8, q^2 values higher than 0.5, r_{pred}^2 values higher than 0.6, and SEP values of 0.412 and 0.267, respectively. The errors between predicted values and experimental values were small, and all data were concentrated near the trend line. Slopes were close to 1.000, and the coefficients of determination (R^2) were both higher than 0.9, which was in line with the internal and external inspection standards of the model. These results showed that the models had good stability and predictive ability. From the contribution rates of the molecular fields, the main factor affecting the logBCF values of the aromatic compounds in both models was the electrostatic field. However, the molecular force field of the CoMFA model only included the steric and electrostatic fields. Analysis of the influence of molecular force fields on the compounds and modification of the molecules has some limitations. The CoMSIA model included five molecular force fields (steric, electrostatic, hydrophobic, hydrogen bond donor, and hydrogen bond acceptor fields) that could more comprehensively reflect the influences of the molecular fields around the compounds.

2.1.4. Analysis of regional substitution sites based on CoMSIA model contour maps

Using PeCB as an example, the contour map (Figures 3 and 4) of the CoMSIA model was analyzed. In the steric field, the green region showed that bulky groups could increase the molecular activity, while the yellow region showed that bulky groups could reduce the molecular activity. In the electrostatic field, the blue region showed

that positive charge groups could increase the molecular activity, while the red region showed that negative charge groups could increase the molecular activity. In the hydrophobic field, hydrophilic groups introduced into the white region could increase the molecular activity, while hydrophobic groups introduced into the yellow region could increase the molecular activity. In the hydrogen bond donor field, the blue-green region showed that adding hydrogen bond donors increased the molecular activity, while the purple region showed that adding hydrogen bond donors did not improve the molecular activity. In the hydrogen bond acceptor field, the magenta region showed that hydrogen bond receptors could increase the molecular activity, while the red region showed that hydrogen bond receptors did not improve molecular activity.

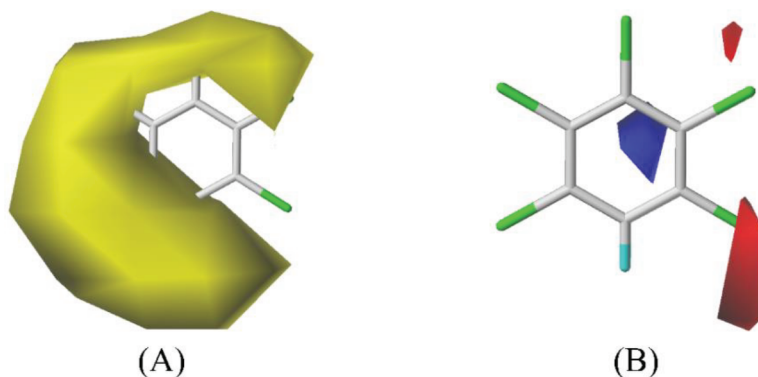


Figure 3. Contour maps of CoMSIA model: steric fields (A), electrostatic fields (B).

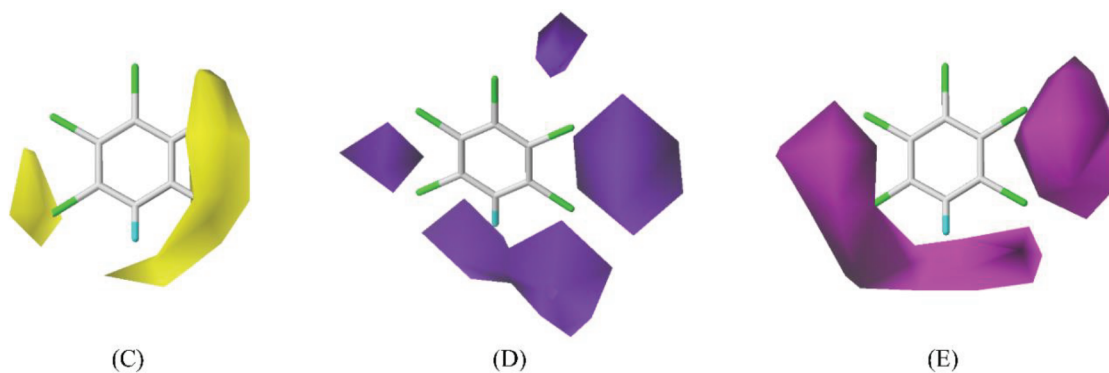


Figure 4. Contour maps of CoMSIA model: hydrophobic field (C), hydrogen-bond donor field (D), hydrogen-bond acceptor field (E).

In the CoMSIA model, the contour map of the steric field showed that the introduction of bulky groups at the 1-, 2-, 3-, 4-, and 6-positions reduced the logBCF values of aromatic compounds. In the electrostatic field contour map, red regions were located near the 5- and 6-positions, which indicated that the introduction of negatively charged groups at the 5- or 6-positions increased the logBCF values of aromatic compounds. The hydrophobic field contour map showed yellow regions around the 3-, 4-, 5-, and 6-positions, which indicated that the introduction of hydrophobic groups at these positions would increase the logBCF values. The hydrogen bond donor field contour map showed purple regions around the 2-, 4-, 5-, and 6-positions, which indicated that the introduction of hydrogen bond donors at these four sites would reduce the logBCF values of the molecules. The hydrogen bond acceptor contour map showed magenta regions around the 2-, 3-, 4-, 5-, and 6-positions, which indicated that introducing hydrogen bond receptors at these positions would increase the logBCF values.

2.2. Determination of substitution site based on coupling of 3D-QSAR and HQSAR models

The contour maps of 3D-QSAR showed only the range of the molecular force fields and could not explain the activity of a specific site. Therefore, a molecular HQSAR method was needed that could generate the molecular holographic descriptors automatically using software. Molecular fragment analysis could show the active contribution of each part of the molecule and solve the problem that active conformations are hard to determine. The analysis results of the HQSAR model constructed using different fragment parameters and fragment sizes are shown in Table 4. From these data, the parameter combinations of fragments A/B/C and A/B/C/Ch, A/B/C/H, and A/B/C/H/Ch gave the same results for the 44 aromatic compounds that did not contain chiral atoms. However, the results of A/B/C/DA and A/B/C/H/DA, A/B/C/Ch/DA, and A/B/C/H/Ch/DA were not the same because polychlorinated phenols contain hydroxyl groups as hydrogen bond donors that could change the molecular chirality. Sixteen groups of models were compared, affording r^2 and q^2 values all higher than 0.9, which indicated that the statistical stability and predictive power of each model was good.²² After careful consideration, the 12th HQSAR model, constructed using a fragment size of 1–7 and fragmentation parameter combination of A/B/C/H/DA, was applied for active site analysis and prediction of the 44 substances (Table 5).

Table 4. HQSAR models of different fragment parameters and sizes.

Model	Fragment distinction	Fragment size	r^2	q^2	SE	SE_{cv}	n	HL
1	A + B + C	4–7	0.984	0.964	0.133	0.243	5	83
2		1–7	0.979	0.945	0.150	0.245	5	97
3	A + B + C + H	4–7	0.983	0.952	0.137	0.230	5	53
4		1–7	0.988	0.948	0.117	0.241	6	71
5	A + B + C + Ch	4–7	0.984	0.964	0.133	0.243	5	83
6		1–7	0.979	0.945	0.150	0.245	5	97
7	A + B + C + D&A	4–7	0.969	0.940	0.186	0.257	5	71
8		1–7	0.981	0.947	0.147	0.244	6	71
9	A + B + C + H + Ch	4–7	0.983	0.952	0.137	0.230	5	53
10		1–7	0.988	0.948	0.117	0.241	6	71
11	A + B + C + H + D&A	4–7	0.987	0.962	0.122	0.206	6	71
12		1–7	0.988	0.965	0.118	0.199	6	71
13	A + B + C + Ch + D&A	4–7	0.982	0.945	0.142	0.249	6	307
14		1–7	0.982	0.947	0.144	0.245	6	307
15	A + B + C + H + Ch + D&A	4–7	0.987	0.959	0.123	0.216	6	71
16		1–7	0.987	0.957	0.119	0.219	6	71

Model 12 was the optimal model. Using ‘Show Contributions’ under the HQSAR module, active site analysis was performed on Model 12. PeCB was used as the target molecule and the active site contribution is shown in Figure 5. A change from red to green represented a change in activity of a group or atom in a molecule. Red, red-orange, and orange indicated an unfavorable effect on the molecular activity, while yellow, blue-green, and green indicated positive effects on molecular activity and white indicated medium activity contribution. The 1-, 5-, and 6-positions of PeCB made positive active contributions and were favorable sites for bioconcentration. By combining the 3D-QSAR electrostatic field contour map, the 5-position was determined to be the most effective replacement site (Figure 6).

Table 5. Predicted logBCF values of aromatic compounds using the HQSAR model.

No.	Compounds	Obs.	Pred.	Relative error (%)
1	1,2-Dichlorobenzene	3.12	3.11842	-0.05%
2	1,3-Dichlorobenzene	3.05	3.24615	6.43%
3	1,4-Dichlorobenzene	3.24	3.29504	1.70%
4	1,2,3-Trichlorobenzene	3.95	3.83305	-2.96%
5	1,2,4-Trichlorobenzene	3.84	3.90013	1.57%
6	1,3,5-Trichlorobenzene	4.05	3.97365	-1.89%
7	1,2,3,4-Tetrachlorobenzene	4.31	4.50982	4.64%
8	1,2,4,5-Tetrachlorobenzene	4.67	4.46736	-4.34%
9	1,2,3,5-Tetrachlorobenzene	4.81	4.52269	-5.97%
10	Pentachlorobenzene	5.02	5.09983	1.59%
11	Hexachlorobenzene	5.64	5.70507	1.15%
12	2-Chlorophenol	2.54	2.36437	-6.91%
13	3-Chlorophenol	2.31	2.51076	8.69%
14	4-Nitrophenol	1.84	1.7368	-5.61%
15	3,5-Dimethylphenol	2.47	2.42848	-1.68%
16	2,3-Dimethylphenol	2.66	2.65927	-0.03%
17	3,4-Dimethylphenol	2.65	2.69881	1.84%
18	2,4,6-Trichlorophenol	3.51	3.6487	3.95%
19	2,4,5-Trichlorophenol	3.67	3.69486	0.68%
20	3,4,5-Trichlorophenol	3.86	3.75793	-2.64%
21	2,3,4,5-Tetrachlorophenol	4.23	4.20688	-0.55%
22	Pentachlorophenol	4.63	4.58074	-1.06%
23	Nitrobenzene	1.47	1.54249	4.93%
24	o-Chloronitrobenzene	2.29	2.22952	-2.64%
25	m-Chloronitrobenzene	2.42	2.3595	-2.50%
26	p-Chloronitrobenzene	2.46	2.43731	-0.92%
27	2,3-Dichloronitrobenzene	3.01	3.01833	0.28%
28	2,4-Dichloronitrobenzene	3.02	3.0667	1.55%
29	2,5-Dichloronitrobenzene	2.92	3.01096	3.12%
30	3,5-Dichloronitrobenzene	3.01	3.07912	2.30%
31	o-Nitrotoluene	2.28	2.24604	-1.49%
32	m-Nitrotoluene	2.31	2.34076	1.33%
33	p-Nitrotoluene	2.37	2.25939	-4.67%
34	4-Chloro-2-nitrotoluene	3.02	2.96993	-1.66%
35	2-Chloro-6-nitrotoluene	3.09	3.09363	0.12%
36	p-Nitroaniline	1.92	1.97179	2.70%
37	o-Dinitrobenzene	1.82	1.8173	-0.15%
38	m-Dinitrobenzene	1.87	1.99902	6.90%
39	p-Dinitrobenzene	1.7	1.555	-8.53%
40	2,4-Dinitrotoluene	2.31	2.49866	8.17%
41	2,6-Dinitrotoluene	2.44	2.29718	-5.85%
42	2,3-Dinitrotoluene	2.49	2.51518	1.01%
43	3,4-Dinitrotoluene	2.32	2.42476	4.52%
44	1,3,5-Trinitrobenzene	2.95	2.86859	-2.76%

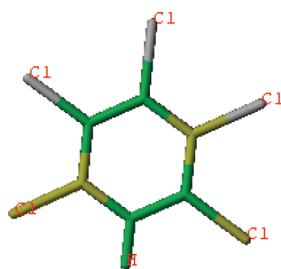


Figure 5. Activity contribution map of HQSAR model.

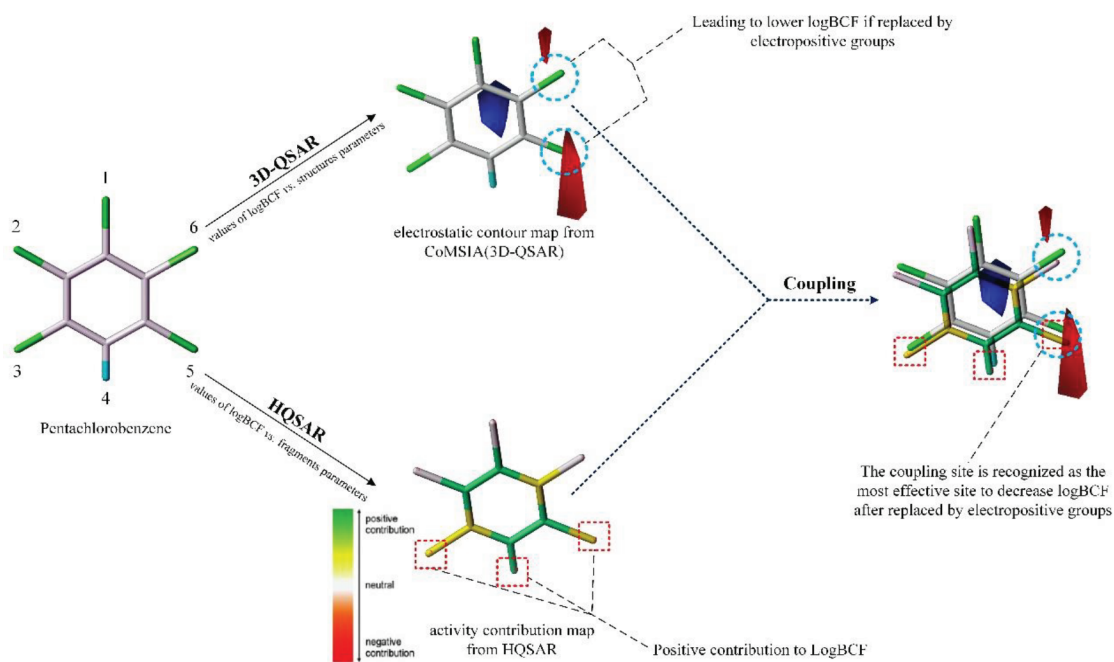


Figure 6. The coupling of 3D-QSAR and HQSAR models.

Table 6. Predicted $\log\text{BCF}$, $\log\text{EC}_{50}$, and $\log\text{K}_{oa}$ values of PeCB derivatives from CoMSIA models.

No.	Substituent groups	$\log\text{BCF}$		$\log\text{EC}_{50}$		$\log\text{K}_{oa}$	
		Pred.	Change rate	Pred.	Change rate	Pred.	Change rate
1	-OH	4.307	-14.20%	5.407	0.76%	6.275	-3.31%
2	-CH ₂ OH	3.649	-27.31%	5.360	-0.11%	6.146	-5.30%
3	-NH ₂	3.643	-27.43%	5.190	-4.79%	5.853	-9.82%
4	-CH ₂ NH ₂	4.864	-3.11%	4.996	-6.89%	5.596	-13.78%
5	-COOH	4.871	-2.97%	5.909	10.12%	7.367	13.51%
6	-NO ₂	4.337	-13.61%	5.816	8.39%	6.855	5.62%
7	-CH ₃	4.395	-12.45%	4.982	-7.16%	5.851	-9.85%

2.3. Evaluation of POP characteristics and functionalities of PeCB derivatives

2.3.1. Evaluation of bioconcentration, toxicity, and migration of PeCB derivatives

Using PeCB as the target molecule, the electrostatic field contour map of the CoMSIA model was compared with the activity contribution map of the HQSAR model, which showed that the 5-position was the target site. Therefore, modification of the 5-position of PeCB was the best method for low-bioconcentration design. Replacing the 5-position atom with seven different groups ($-\text{OH}$, $-\text{CH}_2\text{OH}$, $-\text{NH}_2$, $-\text{CH}_2\text{NH}_2$, $-\text{COOH}$, $-\text{NO}_2$, and $-\text{CH}_3$) with electronegativities lower than that of Cl could lead to different degrees of reduction in bioconcentration. Predictions of bioconcentration, toxicity, and migration of the seven derivatives are shown in Table 6.

The logBCF values of the modified molecules were lower than that of PeCB, which indicated that the introduction of positive charge groups at the 5-position of PeCB could reduce the logBCF value. The decreased range of the logBCF values was 2.97%–27.43%. The BCF values of 5- CH_2OH -PeCB and 5- NH_2 -PeCB were both lower than 5000, which indicated that these two derivatives had weak bioenrichment effects.²³ Using the logEC₅₀ values of aquatic luminescent bacteria,²⁴ a 3D-QSAR model was established to predict the derivatives in this study. Compared with PeCB, the results showed that the toxicity range was -7.16% to 10.12%, which showed that the toxicity of the derivatives was essentially unchanged. Using the logK_{oa} values obtained from the EPI suite database, a 3D-QSAR model was established to predict the derivatives. The results showed that the change rates of the logK_{oa} values of the derivatives were -13.78% to 13.51%, which indicated that the migration abilities were not significantly changed.

2.3.2. Molecular docking analysis of the degradation of derivatives with one degradation enzyme

Due to the lack of data on degradability, the molecular docking method was used to evaluate degradability (total scores) of PeCB and the derivatives. Hydroxylation is an important metabolic pathway for aromatics, and toluene-4-monooxygenase can hydroxylate benzene rings.²⁵ Therefore, the degradation enzyme was selected to dock with PeCB and the derivatives. The total score indicated the binding force; the higher the total score, the better the binding stability.²⁶

The derivatives and target molecule (PeCB) were docked with the degradation enzyme (toluene-4-monooxygenase) using Sybyl. The results (see Figure 7 and Table 7) were used to screen for derivatives with good degradation.

The rationality of the molecular docking results needed to be judged based on the magnitude of the root mean square deviation (RMSD). Molecular docking RMSD values of PeCB and derivatives with toluene-4-monooxygenase were 0.097 nm (PeCB), 0.14 nm ($-\text{NO}_2$), 0.074 nm ($-\text{CH}_3$), 0.069 nm ($-\text{NH}_2$), 0.049 nm ($-\text{COOH}$), 0.043 nm ($-\text{OH}$), 0.058 nm ($-\text{CH}_2\text{OH}$), and 0.045 nm ($-\text{CH}_2\text{NH}_2$) respectively, which satisfied the rationality of molecular docking (≤ 0.2 nm).²⁷

With a nitro ($-\text{NO}_2$) substituent, the total score value was -0.5158, which was lower than that of the target molecule (PeCB). The derivatives with $-\text{CH}_3$, $-\text{NH}_2$, $-\text{COOH}$, $-\text{OH}$, $-\text{CH}_2\text{OH}$, and $-\text{CH}_2\text{NH}_2$ substituents had higher total scores than PeCB, and the increased range of the total score was 270.74% to 603.85%. In addition to 5- NO_2 -PeCB, the other six derivatives had better degradability due to better binding abilities with toluene-4-monooxygenase degradation enzyme compared with the target molecule.

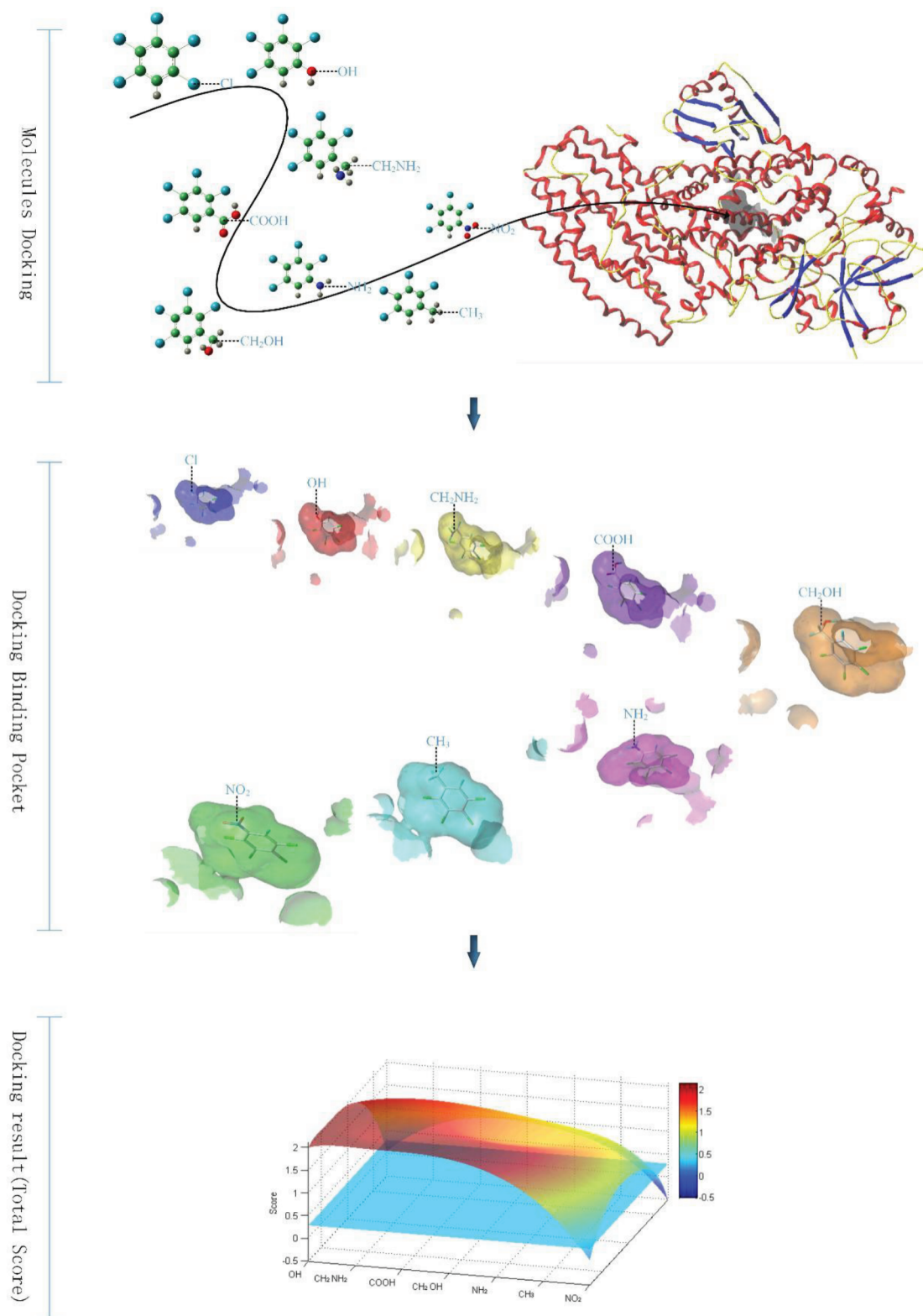
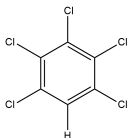
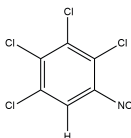
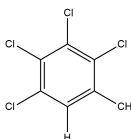
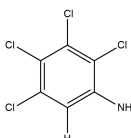
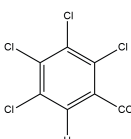
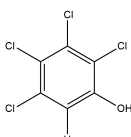
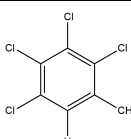
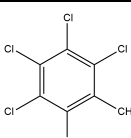


Figure 7. Molecular docking analysis of the derivatives with one degradation enzyme.

Table 7. Docking results (total score) for the derivatives and PeCB with Toluene-4-monoxygenase degradation enzyme.

Compounds	Total score
 PeCB	0.2833
 5-NO ₂ -PeCB	-0.5158
 5-CH ₃ -PeCB	1.0503
 5-NH ₂ -PeCB	1.2193
 5-COOH-PeCB	1.9141
 5-OH-PeCB	1.9940
 5-CH ₂ OH-PeCB	1.4839
 5-CH ₂ NH ₂ -PeCB	1.9650

2.3.3. Evaluation of insulation and flame retardancy of PeCB derivatives

To verify the changes in the flame retardancy and insulation of the derivatives, the dissociation enthalpy and energy gap values of PeCB and its derivatives were calculated using Gaussian 09. A low dissociation enthalpy value indicated good flame retardancy, while a high energy gap indicated good insulation. Geometric optimizations of all compounds were performed using density functional theory at the b3pw91/6-31g* level. The results are shown in Table 8.

Compared with the target molecule, PeCB, the change ranges of the dissociation enthalpies of the derivatives were -7.84% to 3.11%, which indicated no significant changes in the flame retardancies of the derivatives. In addition to 5-NO₂-PeCB, the change ranges of the energy gap of the derivatives compared with that of PeCB were -10.71% to 2.25%, which indicated that the insulating properties of the derivatives were essentially unchanged.

2.4. Conclusions

The results of the CoMFA and CoMSIA models showed cross-validation coefficients (q^2) of 0.694 and 0.869, non-cross-validation coefficients (R^2) of 0.913 and 0.984, standard errors of cross-validation of 0.412 and 0.267, and interaction test coefficients (r^2_{pred}) of 0.761 and 0.900, respectively. The results indicated that the established model had strong stability and good prediction ability, with the electrostatic field found to be the main factor affecting the logBCF values of aromatic compounds.

According to the contour maps of the electrostatic field, the introduction of positive groups at the 3- and 5-positions could reduce the bioconcentration of the aromatic compounds. Based on the electrostatic field contour map and the HQSAR activity contribution map, seven novel molecules with low logBCF values were designed and no significant changes in toxicity and mobility were observed. Except for 5-NO₂-PeCB molecules, the binding abilities of the modified molecules with toluene-4-monooxygenase degradation enzyme were increased. Except for 5-NO₂-PeCB, no significant changes in insulation were observed. The flame retardancy of seven kinds of modified molecules remained essentially unchanged.

3. Experimental

3.1. Methods

Herein, 3D-QSAR and HQSAR models were calculated using Sybyl molecular simulation software. Default values were used for all parameters except when specified in the calculation process.

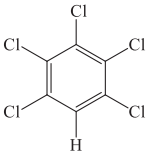
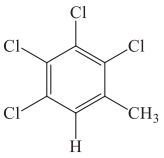
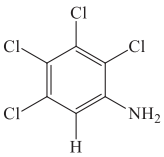
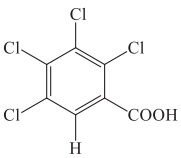
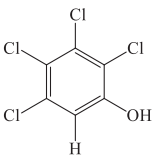
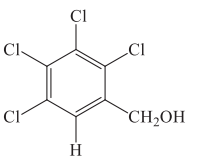
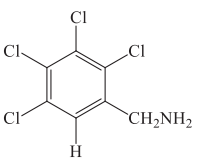
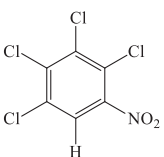
3.2. Establishment of monophenyl aromatic compound bioconcentrations using 3D-QSAR model

The data source of 3D-QSAR for aromatic compounds was constructed using the logBCFs of 44 monophenyl aromatic compounds as the dependent variables and the structure parameters as the independent variables.²⁴ During the model construction process, in accordance with the 4:1 ratio, 36 aromatic molecules were selected stochastically as training sets to establish a 3D-QSAR model with eight other aromatic molecules used as test sets.

3.2.1. Molecular structure modeling and alignment of aromatic compounds

Subsequent calculations using Sybyl software used the molecular structure adjusted to a stable conformation. We selected 36 training set molecules as the research object and used the Powell conjugate gradient method

Table 8. Energy gap and C-Cl bond dissociation enthalpy results for PeCB derivatives.

Compounds	C-Cl bond dissociation enthalpy (kcal/mol)	Energy gap (a.u.)
	86.4996	0.2091
	87.1428	0.2133
	89.1910	0.1867
	79.7195	0.1931
	88.3225	0.2002
	88.7988	0.2138
	88.3451	0.2131
	82.0375	0.1698

to optimize 36 single-benzene aromatic molecules with molecular mechanics using the Minimize module with Gasteiger–Hückel charge. Consequently, the low-energy conformation of each molecule was obtained as a stable conformation. Optimized molecules were then stored in the database to be aligned using the benzene ring as the common skeleton and the hexachlorobenzene molecule with the largest logBCF value as the template molecule. For each compound in the Tripos field, the maximum number of optimizations was 10,000 and the energy convergence was limited to 0.001 kJ/mol.

3.2.2. CoMFA and CoMSIA analyses of aromatic compound bioconcentrations

CoMFA and CoMSIA analyses were completed by QSAR module calculation in Sybyl software. The CoMFA force field was calculated to obtain the molecular force field parameters. The molecular force field types were electrostatic (E) and stereoscopic (S) fields, and the field energy threshold value was the default value of 30 kcal mol⁻¹. Meanwhile, the logBCF values for 36 aromatic compounds were input into the Training table and CoMFA field parameters were calculated automatically using Sybyl via Autofill. Partial least squares (PLS) analysis was then used to establish the relationship between the target compound structure and biological activity. In PLS analysis, cross-validation of the leave-one-out method used for the training set compounds was performed to obtain the cross-validation coefficient (q^2) and the best number of principal components (n). Next, regression analysis was performed using nonvalidation regression (no validation) and the non-cross-validation coefficient (r^2), standard error of estimate (SEE), and test value (F) were obtained to complete the establishment of the CoMFA model. Compared with the CoMFA model, CoMSIA can determine the structure–activity relationships of compounds more intuitively owing to the added hydrophobic (H), hydrogen bond acceptor (A), and hydrogen bond donor (D) fields. Furthermore, more meta probe atoms were used for detection and the results were influenced relatively little due to the compound matching rule. Although CoMSIA can overcome some of the inherent defects of CoMFA, it may not be able to obtain better results.²⁸ This study will use the CoMFA and CoMSIA methods in 3D-QSAR studies to verify and supplement each other.

3.3. HQSAR model of aromatic compound bioconcentrations

Calculation of the HQSAR model was completed using the Minimize and HQSAR modules of Sybyl software. First, we selected 44 single-benzene aromatic molecules as the target objects and the Powell energy gradient method was used to optimize these molecules with molecular mechanics in the Minimize module with Gasteiger–Hückel charge. For each compound in the Tripos field, the maximum number of optimizations was 10,000 and the energy convergence was limited to 0.005 kJ/mol. The HQSAR model can divide molecules into different structural fragments, with substructure types represented by these fragments including atomic type (A), chemical bond type (B), connectivity (C), hydrogen atom (H), chirality (Ch), and proton donor and acceptor (D&A). A fragments can distinguish different atomic species, B can distinguish differences in chemical bonds among atoms, C can identify atomic hybridization states, Ch can identify atom chirality, and D&A can distinguish whether hydrogen bonds are present.⁹ Herein, additional fragment parameters were added to the A/B/C fragments to obtain eight fragment parameter combinations (A/B/C, A/B/C/H, A/B/C/Ch, A/B/C/DA, A/B/C/H/Ch, A/B/C/H/DA, A/B/C/Ch/DA, and A/B/C/H/Ch/DA), By selecting fragment sizes of 4–7 and 1–7 and 12 hologram lengths (53, 59, 61, 71, 83, 97, 151, 199, 257, 307, 353, and 401), we established 16 groups of HQSAR models.

Acknowledgments

This research was supported by the Key Projects in the National Science & Technology Pillar Program in the Eleventh Five-Year Plan Period (No. 2008BAC43B01) and supported by the Fundamental Research Funds for the Central Universities 2017XS058.

We thank Simon Partridge, PhD, from Liwen Bianji, Edanz Editing China, for editing the English text of a draft of this manuscript.

References

1. Sari, A. A.; Yasin, H.; Tachibana, S.; Hadibarata, T. *Water Air Soil Poll.* **2016**, *9*, 227-317.
2. Yang, Y.; Yu, G.; Deng, S.; Wang, S.; Xu, Z.; Huang, J.; Wang, B. *Chem. Eng. J.* **2012**, *192*, 284-291.
3. Roscales, J. L.; Mu noz-Arnanz, J.; Morales, L.; Abad, E.; Jimenez, B. In *Halogenated Persistent Organic Pollutants 2013, Proceedings of the 33th International Symposium*; Daegu, South Korea, August 2013, pp. 137-140.
4. Liu, G.; Zheng, M.; Cai, Z. *Environ. Sci. Pollut. R.* **2014**, *21*, 13656-13663.
5. Karich, A.; Ullrich, R.; Scheibner, K.; Hofrichter, M. *Front. Microbiol.* **2017**, *8*, 1463.
6. Wang, P. MSc, University of Science and Technology of China, Hefei, China, 2015.
7. Li, F.; Zhao, G. *Food. Ferment. Ind.* **2014**, *40*, 128-134.
8. Zhao, Y.; Wang, W.; Liu, H.; Huang, L.; Lu, B.; Yang, G.; Fang, L. *China Measurement & Testing Technology* **2016**, *42*, 35-38.
9. Jones, K. C.; De, V. P. *Environ. Pollut.* **1999**, *100*, 209-221.
10. Liu, J. In *Encyclopedia of Toxicology (Third Edition)*; Elsevier: Amsterdam, the Neterlands, 2014, pp. 773-774.
11. Treves, K.; Shragina, L.; Rudich, Y. *Atmos. Environ.* **2001**, *35*, 5843-5854.
12. Arnot, J. A.; Gobas, F. A. *Environ. Rev.* **2006**, *14*, 257-297.
13. Weisglas-Kuperus, N.; Vreugdenhil, H. J.; Mulder, P. G. *Toxicol. Lett.* **2004**, *149*, 281-285.
14. Loomis, D. *Food. Chem. Toxicol.* **1997**, *35*, 1230.
15. Fernández, P.; Bayona, J. M. *J. Sep. Sci.* **2015**, *12*, 802-806.
16. Togunde, O. P. PhD, University of Waterloo, Waterloo, Canada, 2012.
17. Torreiro-Melo, A. G.; Silva, J. S.; Bianchini, A.; Zanardilamardo, E.; Carvalho, P. S. *Chemosphere* **2015**, *132*, 17-23.
18. Cherkasov, A.; Muratov, E. N.; Fourches, D.; Varnek, A.; Baskin, L. L.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y. C.; Todeschini, R. et al. *J. Med. Chem.* **2014**, *57*, 4977-5010.
19. Cichero, E.; Fossa, P. *J. Mol. Model.* **2012**, *18*, 1573-1582.
20. Sharma, M. C.; Sharma, S.; Sahu, N. K.; Kohli, D. V. *J. Saudi. Chem. Soc.* **2013**, *17*, 167-176.
21. Golbraikh, A.; Tropsha, A. *Mol. Graph. Mod.* **2002**, *20*, 269-276.
22. Clark, M.; Iii, R. D. C.; Jones, D. M.; Patterson, D. E.; Simeroth, P. E. *Tetrahedron Comput. Methodol.* **1990**, *3*, 47-59.
23. Koleva, Y.; Georgieva, M.; Zlatkov, A.; Georgieva, S.; Mitkov, J. *Oxid. Commun.* **2013**, *36*, 1135-1142.
24. Zhou, G. MSc, Lanzhou University, Lanzhou, China, 2009.
25. Mitchell, K. H.; Rogge, C. E.; Gierahn, T.; Gierahn, T.; Fox, B. G. *P. Natl. Acad. Sci. USA* **2003**, *100*, 3784-3789.
26. Holt, P. A.; Chairs, J. B.; Trent, J. O. *J. Chem. Inf. Model.* **2008**, *39*, 1602-1615.
27. Zheng, J.; Zhao, B.; Yan, H.; Zhang, K.; Zhang, D.; Zhao, S. *Acta Chim. Sinica* **2011**, *69*, 2026-2030.
28. Chen, Y.; Cai, X.; Jiang, L.; Li, Y. *Ecotox. Environ. Safe.* **2016**, *124*, 202-212.
29. Sun, S.; Mi, S.; You, J.; Yu, J.; Hu, S.; Liu, X. *Acta. Phys-Chim. Sin.* **2013**, *29*, 2402-2409.