QSARS FOR PREDICTION OF THE TOXICITY OF PERSISTENT ORGANIC POLLUTANTS USING QUANTUM CHEMICAL DESCRIPTORS

Shuqiong Shi, Lili Wang, Junfeng Niu

State Key Laboratory of Water Environment Simulation, School of Environment, Beijing Normal University, Beijing 100875, P.R. China

Abstract

Quantitative structure–activity relationship (QSAR) models were developed for the toxicity of polychlorinated dibenzo-*p*-dioxins (PCDDs), dibenzofurans (PCDFs), biphenyls (PCBs) and naphthalenes (NAPs), respectively, using partial least square (PLS) regression. Quantum chemical descriptors computed by semi-empirical PM3 method were used as predictor variables. Four optimal QSAR models are developed for 25 PCDDs, 35 PCDFs, 25 PCDDs and 35 PCDFs together, 30 PCBs and 5 NAPs together, respectively. All the cross-validated Q^2_{cum} values of the four QSAR models are higher than 0.50, which shows that these models have good predictive capabilities for the biological toxicity of these persistent organic pollutants (POPs). The results of this study provide a rapid, simple and valid means of predicting the toxicity of POPs from the chemical structure.

Introduction

Persistent organic pollutants (POPs) such as organochlorine compounds (OCs), polychlorinated dibenzo-*p*-dioxins (PCDDs), dibenzofurans (PCDFs), and biphenyls (PCBs) have been detected in many kinds of environmental matrices even in remote areas such as the Arctic[1-3]. Because of their ubiquitous distribution, toxicity, persistence and bioaccumulation potential in the food web, POPs have raised concern about their adverse effects such as carcinogenicity, teratogenicity, and mutagenicity on organisms and human[4-6]. However, because of high cost, time-consuming process, limits of detection and lack of adequate standard materials, toxicity data of POPs are rather scarce. To solve these problems, quantitative structure-activity relationship (QSAR) models, which correlate and predict toxicity data of POPs from their molecular structural descriptors, provide valuable approach in research into the toxicity without any experiments[7-9]. POPs cause adverse biological effects after binding to a common intracellular cytosolic protein called the aryl hydrocarbon receptor (AhR), so the key point in predicting the toxic effects of POPs is the estimation of their binding to AhR[10]. In this study, quantum chemical descriptors were used to develop reliable QSAR models for the binding affinities (BA) of POPs for aryl hydrocarbon receptors (AhRs) using partial least square (PLS) algorithm.

Materials and methods

The experimental toxicity parameter, BA of POPs for AhRs in *vitro* rat hepatocyte assays reported, was adopted in this study[11-13]. These toxicity values were collected by Ashek et al.[14]. In previous reports, we found that models computed by PM3 Hamiltonian suggested a better predictive capability for the toxicity of hydroxylated and quinoid PCB metabolites than models developed using other semi-empirical methods such as AM1 and MNDO[15]. Therefore, molecular structural descriptors were calculated for PCDDs, PCDFs, PCBs and NAPs by

semi-empirical PM3 methods in the present study. All calculations were performed using MOPAC (2000) contained in the CS Chem3D Ultra (Version. 6.0). A total of 23 MOPAC-derived descriptors that reflect the overall characters of the chemicals were computed using PM3 method in this study. A full list is given in Table 1.

Table 1 List of molecular structural descriptors of POPs

Symbols	Description
M_W	Molecular weight
ΔH_f	Standard heat of formation (kcal)
TE	Total energy
EE	Electronic energy
CCR	Core-core repulsion energy
E_{HOMO}	The energy of the highest occupied molecular orbital
E_{HOMO-1}	The energy of the second highest occupied molecular orbital
E_{LUMO}	The energy of the lowest unoccupied molecular orbital
E_{LUMO+1}	The energy of the second lowest unoccupied molecular orbital
q_{CL}^+	The largest positive atomic charge on a chlorine atom
q_H^+	The most positive net atomic charges on a hydrogen atom
q_C^-	The largest negative atomic charge on a carbon atom
q_O^-	The largest negative atomic charge on a oxygen atom
μ	Dipole moment
μ_x	X-axis dipole moment
μ_y	Y-axis dipole moment
μ_z	Z-axis dipole moment
α	Average molecular polarizability
IP	Ionization potential

The Simca (Simca-S Version 6.0, Umetri AB and Erisoft AB) software was used to perform the PLS analysis. The criterion used to determine the model dimensionality is cross validation (CV). When the cumulative cross-validated regression coefficient (Q^2) for the extracted components, Q^2_{cum}, is larger than 0.5, the model is considered to have a good predictive ability. Model adequacy was mainly characterized by the number of observations used for model building in the training set, the number of PLS principal components (*A*), Q^2_{cum}, the correlation coefficient between observed and fitted values (*R*), the general standard error (*SE*) and the significance level (*p*)[16].

Results and discussion

In a PLS model, variable importance in the projection (*VIP*) is a parameter in the PLS analysis that shows the importance of a variable in a PLS model. PLS analysis with log1/EC_{50} values of POPs as dependent variable and the 23 quantum chemical descriptors as independent variables generate many results. The optimal model, which has the largest Q^2_{cum}, largest *R* and smallest *p*, was obtained through stepwise culling the model with the smallest *VIP* value.

The 95 POPs are divided into four groups for developing significant QSAR models depending on their parent molecules. Following the procedure

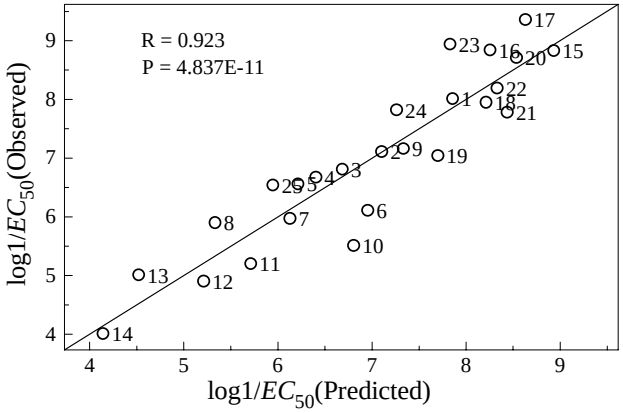

Figure 1 Plots of observed vs. predicted log1/EC_{50} values of PCDDs in model (1)

described above, four models were obtained using computed molecular descriptors by semi-empirical PM3 method for log1/EC_{50} values of the POPs. For example, models (1) to (4) were developed for 25 PCDDs, 35 PCDFs, combination of 25 PCDDs and 35 PCDFs, and 30 PCBs and 5 NAPs together, respectively. Based on the unscaled pseudo-regression coefficients of the independent variables and constants transformed from PLS results, analytical QSAR equations from models (1)~(4) were obtained and shown in Eq. 1 to 4:

Model (1) for PCDDs:

$$\log 1/EC_{50} = 3.281 - 6.588\times10^{-1}(E_{LUMO}-E_{HOMO})^2 - 7.377\times10^{-4}\ CCR - 2.465\ E_{HOMO} + 1.109\times10^{-2}\ M_W - 3.648\ (E_{LUMO}+E_{HOMO}) + 4.370\ E_{LUMO+1} \quad (1)$$

n=25, A=3, $R^2_{X(adj.)(cum)}$=0.989, $R^2_{Y(adj.)(cum)}$=0.853, Eig=0.363, Q^2_{cum}=0.816, R=0.923, SE=0.593

Model (2) for PCDFs:

$$\log 1/EC_{50} = -46.961 - 1.343\times10^{-1}(E_{LUMO} - E_{HOMO})^2 - 2.151(E_{LUMO} - E_{HOMO}) - 22.439\ q_C^- + 38.800\ q_H^+ - 6.590\times10^{-1}\mu_y - 7.701\ E_{HOMO} - 3.38\times10^{-1}\mu_x \quad (2)$$

n=35, A=2, $R^2_{X(adj.)(cum)}$=0.608, $R^2_{Y(adj.)(cum)}$=0.707, Eig=1.089, Q^2_{cum}=0.629, R=0.841, SE=0.769

Model (3) for PCDD/Fs:

$$\log 1/EC_{50} = -27.89 - 3.269E_{HOMO-1} + 3.325\times10^{-2}\alpha - 2.793\times10^{-1}\mu_x + 24.622q_H^+ - 7.680\times10^{-2}(E_{LUMO}-E_{HOMO})^2 \quad (3)$$

n=60, A=2, $R^2_{X(adj.)(cum)}$= 0.549, $R^2_{Y(adj.)(cum)}$= 0.686, Eig=1.140, Q^2_{cum}=0.603, R=0.829, SE=0.870

Model (4) for PCBs and NAPs:

$$\log 1/EC_{50} = -14.721 + 5.545\times10^{-3}M_W - 2.209E_{HOMO-1} - 7.872\times10^{-1}E_{LUMO} - 24.284\ q_H^+ + 1.632\times10^{-1}\mu_y - 4.378\times10^{-3}(E_{LUMO} - E_{HOMO})^2 + 6.324\times10^{-1}\ q_C^- \quad (4)$$

n=35, A=2, $R^2_{X(adj.)(cum)}$=0.616, $R^2_{Y(adj.)(cum)}$=0.782, Eig=1.532, Q^2_{cum}=0.734, R=0.884, SE=0.499

Table 2 The *VIP* values for the molecular structural descriptors included in models (1)~(4)

Model(1)		Model(2)		Model(3)		Model(4)	
Variables	*VIP*	Variables	*VIP*	Variables	*VIP*	Variables	*VIP*
$(E_{LUMO}-E_{HOMO})^2$	1.591	$(E_{LUMO}-E_{HOMO})^2$	1.349	E_{HOMO-1}	1.338	M_W	1.413
CCR	1.020	$E_{LUMO}-E_{HOMO}$	1.343	α	1.273	E_{HOMO-1}	1.395
E_{HOMO}	0.830	q_C^-	0.994	μ_x	0.753	E_{LUMO}	1.144
M_W	0.801	q_H^+	0.987	q_H^+	0.750	q_H^+	0.923
$E_{LUMO}+E_{HOMO}$	0.756	μ_y	0.748	$(E_{LUMO}-E_{HOMO})^2$	0.679	μ_y	0.637
E_{LUMO+1}	0.726	E_{HOMO}	0.658			$(E_{LUMO}-E_{HOMO})^2$	0.577
		μ_x	0.649			q_C^-	0.397

Models (1)~(4) suggested that molecular structural characteristics affected the toxicity of these chemicals. As all the cross-validated Q^2_{cum} values of models (1)~(4) are larger than 0.50, these models are surely stable and have good prediction capability. It is also obvious that the Q^2_{cum} values of model (1) and model (2) are larger than the Q^2_{cum} values of model (3), the PLS models developed for PCDDs and PCDFs, respectively, are more reliable than that for PCDDs and PCDFs together. This may be reasonable since the parent structures of PCDDs, PCDFs,

PCBs and NAPs are different from each other. In conclusion, these models provided a reliable way to predict the toxicity of POPs from their chemical structures.

Acknowledgements
The research was supported by the National Basic Research Program of P.R. China (973 Project, 2003CB415204, 2002CB412409).

References

1. Bernhoft A, Wiig Ø, Skaare JU. *Environ. Pollut.* 1997; 95:159.
2. Tanabe S, Senthilkumar K, Kannan K, Subramanian AN. *Environ. Contam. Toxicol.* 1998; 34:387.
3. Brunström B, Halldin K. *Toxicol. Lett* . 2000; 112:111.
4. Ross PS, Birnbaum LS. *Hum. Ecol. Risk Assess.* 2003; 9:303.
5. Bayen S, Lee HK, Obbard JP. *Environ. Res*. 2007; 103:375.
6. Bustnes JO, Helberg M, Strann K, Skaare JU. *Environ. Pollut.* 2006; 144:893.
7. Netzeva TI, Schultz TW. *Chemosphere* 2005; 61:1632.
8. Liu Y, Chen JN, Zhao JS, Yu HX, Wang XD, Jiang J, Jin HJ, Zhang JF, Wang LS. *Chemosphere* 2005; 20:791.
9. Juranić IO, Drakulić BJ, Petrović SD, Mijin DZ, Stanković MV. *Chemosphere* 2006; 62:641.
10. Cheney BV, Tolly T. *Int. J. Quantum Chem.* 1979; 16:87.
11. Cheung ENY, McKinney JD. *Mol. Toxicol.* 1989; 2:39.
12. Okey AB. *Pharmacol. Therapeut.* 1990; 45:241.
13. Waller CL, McKinney JD. *J. Med. Chem.* 1992; 35:3660.
14. Ashek A, Lee C, Park H, Cho SJ. *Chemosphere* 2006; 65:521.
15. Niu JF, Long XX, Shi SQ. *J. Mol. Model.* 2007; 13:163.
16. Chen JW, Xue XY, Schramm KW, Quan X, Yang FL, Kettrup A. *Chemosphere* 2002; 48:535.