Bull. Environ. Contam. Toxicol. (2004) 73:659–665

DOI: 10.1007/s00128-004-0477-y

Bulletin of Environmental Contamination and Toxicology

Prediction of the Ability of PAHs to be Photocytotoxic to a Cell Line from the Rainbow Trout (*Oncorhynchus mykiss*) Gill

J. F. Niu, G. Yu

Department of Environmental Science and Engineering, POPs Research Center, Tsinghua University, Beijing 100084, People's Republic of China

Received: 15 November 2003/Accepted: 14 July 2004

Polycyclic aromatic hydrocarbons (PAHs), produced mainly by incomplete combustion of fossil fuels and natural materials (i.e., forest fires, prairie fires, or agricultural burning), are ubiquitous pollutants with high toxicity (Chen et al. 2000 and 2001; Niu et al. 2003). The US-EPA has included 16 PAHs on the list of priority pollutants. Generally, most PAHs occurred in nature are exposed to the irradiation of sunlight. In addition to the photolysis of PAHs under sunlight irradiation (Niu et al. 2003), some studies in recent years showed that quite a few PAHs have photocytotoxicity upon absorbing UV radiation (Schirmer et al. 1997; Schirmer et al. 1998a). The photocytotoxicity of PAHs has been studied in a number of aquatic organisms (Oris and Giesy 1986; Kagan et al. 1987). However, because of large expenditures of money and time, it is difficult to determine the photocytotoxicity for all PAHs. Thus the development of quantitative structure-activity relationship (QSAR) for photocytotoxicity is of great importance for the purpose of risk assessment of these widespread pollutants.

Various molecular structural descriptors, like constitutional descriptors, electrostatic descriptors, topological descriptors, geometrical descriptors and quantum chemical descriptors (Chen et al. 2000; McClelland and Jurs 2000; Yaffe and Cohen 2001; Hu and Aizawa 2003), have been used to develop QSARs or quantitative structure-property relationships (QSPRs). As quantum chemical descriptors can be easily obtained by computation, can clearly describe defined molecular properties, and are not restricted to closely related compounds, the development of QSAR models in which quantum chemical descriptors are used is of great importance. As partial least squares (PLS) regression can analyze data with strongly collinear, noisy and numerous X variables (Wold et al. 2001), it was used for model development in the present study. PLS finds the relationship between a matrix Y (containing dependent variables) and a matrix X (containing predictor variables) by reducing the dimension of the matrix X while concurrently maximizing the correlation between them (Chen et al. 2003).

MATERIALS AND METHODS

Schirmer et al. (1998a) investigated the ability of 16 priority PAHs to be photocytotoxic to a cell line from the rainbow trout (*Oncorhynchus mykiss*) gill,

Correspondence to: G. Yu

RTgill-W1. They found that 13 priority PAHs were photocytotoxic and three PAHs showed no photocytotoxicity. For the PAHs that were identified as photocytotoxic to RTgill-W1 cells, relative photocytotoxic potencies of PAHs were calculated relative to fluoranthene (Schirmer et al. 1998a). Calculations were based on the mean of the photocytotoxic ability (EC_{50}) values for the three assays, i.e., alamar Blue (metabolic activity), 5-carboxyfluorescein diacetate acetoxymethyl ester (CFDA-AM; cell membrane integrity), and neutral red (lysosomal membrane integrity), at 24 h after UV irradiation. The results of EC_{50} values to the PAHs are reproduced in Table 1. The 13 PAHs with EC_{50} values determined constitute the training set of the study. Three PAHs for which show no photocytotoxicity constitute the prediction set.

Table 1. The PAHs under study and their photocytotoxicity*.

No.	Compounds	EC_{50} (nM) (*Obs.*)	$\log(1/EC_{50})$ (*Obs.*)	$\log(1/EC_{50})$ (*Pred.*)	*Diff.*	*SE*
1	Naphthalene			-4.282		±0.380
2	Acenaphthylene	24000	-4.380	-3.746	-0.634	±0.301
3	Acenaphthene	26000	-4.415	-4.403	-0.012	±0.399
4	Fluorene			-4.546		±0.421
5	Phenanthrene	54000	-4.732	-3.669	-1.063	±0.290
6	Anthracene	49	-1.690	-2.290	0.600	±0.186
7	Fluoranthene	93	-1.968	-2.949	0.981	±0.209
8	Pyrene	55	-1.740	-2.086	0.346	±0.193
9	Benzo[a]anthracene	28	-1.447	-2.188	0.741	±0.189
10	Chrysene			-2.629		±0.190
11	Benzo[b]fluoranthene	111	-2.045	-2.561	0.516	±0.187
12	Benzo[k]fluoranthene	184	-2.265	-2.144	-0.121	±0.190
13	Benzo[a]pyrene	11	-1.041	-0.410	-0.631	±0.536
14	Dibenz[a,h]anthracene	132	-2.121	-1.967	-0.154	±0.200
15	Benzo[g,h,i]perylene	12	-1.079	-1.307	0.228	±0.267
16	Indeno[1,2,3-cd]pyrene	176	-2.246	-1.517	-0.729	±0.241

* *Obs.*: Observed values determined by Schirmer et al. (1998a), EC_{50} values were calculated as the mean of the mean EC_{50} values obtained for each fluorescent indicator dye 24 h after irradiation; *Diff.*: difference between observed and predicted $\log(1/EC_{50})$ values; *SE*: Standard errors for the predicted $\log(1/EC_{50})$ values.

The computational time for semi-empirical molecular orbital methods is much shorter than needed by *ab initio* methods. Recently Seward et al. (2002) studied the effect of precision of molecular orbital descriptors on toxicity modeling of selected pyridines. They found instances where calculated quantum chemical descriptors, for example, the energy of the highest occupied molecular orbital (E_{HOMO}), energy of the lowest unoccupied molecular orbital (E_{LUMO}), varied both between different Hamiltonian versions and other similar software packages. However, Seward et al. (2002) found the variability in no way affects the statistical significance of models. For the current study, PM3 (Stewart 1989a,b)

Hamiltonian was applied for computing the quantum chemical descriptors.

MOPAC (2000) contained in the CS Chem3D Ultra (Ver. 6.0) was used to compute quantum chemical descriptors. The molecular structures were optimized using eigenvector following (Baker 1986), a geometry optimization procedure within MOPAC 2000. The geometry optimization criteria GNORM was set at 0.1. A total of 11 MOPAC derived descriptors reflecting the overall character of the PAH molecules were used in this study. These are molecular weight (*Mw*), final heat of formation (*HOF*), total energy (*TE*), electronic energy (*EE*), core-core repulsion energy (*CCR*, $CCR = TE - EE$), average molecular polarizability (α), dipole moment (μ), E_{HOMO}, E_{LUMO}, most positive net atomic charges on a hydrogen atom (q_H^+), and largest negative atomic charge on a carbon atom (q_C^-). The values of the molecular descriptors are listed in Table 2. The compound numbers in Table 2 correspond to those in Table 1. The unit of *HOF* is kilocalorie, and units of energy, charge, dipole and polarizability were electron volts (eV), atomic charge units (a.c.u) and atomic units (a.u.) respectively. In addition, three combinations of frontier molecular orbital energies, E_{LUMO}-E_{HOMO}, $(E_{LUMO}$-$E_{HOMO})^2$ and E_{LUMO}+E_{HOMO} were also selected as independent variables. The E_{LUMO}-E_{HOMO} and E_{LUMO}+E_{HOMO} can be related to absolute hardness and electronegativity respectively (Pearson 1986; Faucon et al. 1999). Furthermore, they were proven to be significant in PAH QSAR (Veith et al. 1995; Boese et al. 1998) and QSPR studies (Chen et al. 2000 and 2001).

Table 2. Selected quantum chemical descriptors of the PAHs.

No.	*Mw*	*HOF*	*TE*	*EE*	α	μ	E_{HOMO}	E_{LUMO}	q_H^+	q_C^-
1	128.173	40.674	-1307.307	-6640.178	83.815	0	-8.835	-0.408	0.105	-0.101
2	152.195	73.129	-1543.181	-8573.164	101.791	0.115	-9.055	-1.062	0.115	-0.121
3	154.211	38.844	-1575.334	-9067.874	98.450	0.221	-8.589	-0.352	0.106	-0.094
4	166.222	48.868	-1693.533	-9871.993	108.829	0.143	-8.842	-0.335	0.107	-0.106
5	178.233	55.026	-1811.913	-10911.542	123.590	0.006	-8.740	-0.535	0.111	-0.102
6	178.233	61.662	-1811.624	-10791.631	131.370	0.004	-8.248	-0.971	0.108	-0.100
7	202.255	80.012	-2048.112	-13135.528	143.475	0.093	-8.725	-1.044	0.107	-0.111
8	202.255	64.141	-2048.430	-13265.392	147.556	0	-8.249	-1.010	0.106	-0.102
9	228.293	74.463	-2316.254	-15640.549	173.150	0.011	-8.328	-0.934	0.114	-0.102
10	228.293	70.854	-2316.410	-15758.755	168.154	0	-8.496	-0.783	0.111	-0.102
11	252.315	93.630	-2552.751	-18268.178	188.425	0.107	-8.663	-1.074	0.109	-0.105
12	252.315	178.311	-2552.652	-18087.710	193.254	0.054	-8.400	-1.012	0.110	-0.107
13	252.315	92.090	-2552.807	-18366.034	218.447	0.020	-7.672	-1.511	0.116	-0.104
14	278.353	87.926	-2820.503	-20950.778	218.125	0	-8.377	-0.918	0.114	-0.102
15	276.337	84.381	-2790.422	-21395.948	208.738	0.022	-8.122	-1.167	0.113	-0.099
16	276.337	104.929	-2789.543	-21024.174	215.575	0.170	-8.235	-1.386	0.108	-0.108

Simca (Simca-S Version 6.0, *Umetri AB and Erisoft AB*) software was used to perform the PLS analysis. The conditions for the computation were based on the

default values of the software. The criterion used to determine the model dimensionality - the number of significant PLS components - is cross validation (CV). With CV, when the fraction of the total variation of the dependent variables that can be predicted by a component, Q^2, for the whole data set is larger than a significance limit (0.097), the tested PLS component is considered significant. When the cumulative Q^2 for the extracted components, Q^2_{cum}, is larger than 0.5, the model is considered to have a good prediction ability. Model adequacy was mainly measured as the number of PLS principal components (k), Q^2_{cum}, the correlation coefficient between observed values and predicted values (R), and the significance level (p).

RESULTS AND DISCUSSION

In PLS models, Variable Importance in the Projection (*VIP*) is a parameter that shows the importance of a variable. According to the manual of Simca-S (Ver. 6.0), *VIP* is the sum over all model dimensions of the contributions of variable influence (*VIN*) for a given PLS dimension (a) and a given X term (k), VIN^2 is computed from the squared PLS weight of that X term, multiplied by the percent explained the residual sum of squares (*SS*) by that PLS dimension. *VIP* value is calculated from the accumulated value over all PLS dimensions,

$$VIP_k = \sum_{a=1}^{A} (VIN)_k^2$$

divided by the total percent explained *SS* by the PLS model and multiplied by the number of terms in the model. Thus terms with large values of *VIP* are the most relevant for explaining dependent variable.

For the PLS analysis process, a PLS model with all the predictor variables was calculated at first. Then the variable with the lowest *VIP* value was eliminated and a new PLS analysis was performed, leading to a new PLS model. This procedure was repeated until the optimal PLS model was selected with respect to the statistics Q^2_{cum}, R, and p.

The above described PLS analysis procedure with $\log(1/EC_{50})$ as dependent variable and the 14 quantum chemical descriptors as independent variables, for the 13 PAHs contained in the training set, led to QSAR model (1) as the optimal one. The concrete results for Model (1) are shown in Table 3. In Table 3, $R^2_{X(adj)(cum)}$ and $R^2_{Y(adj)(cum)}$ stand for cumulative variance of all the X's and Y's, respectively, explained by all extracted components. *Eig* stands for the eigenvalue which denotes the importance of the PLS principal components. So it can be seen from Table 3 that one PLS principal component was selected in Model (1), and the one PLS principal component explained 51.9% of the variance of the independent variables, and 72.3% of the variance of the dependent variable.

Table 3. Model fitting results for Model (1).

k	*Eig*	$R^2_{X(adj)(cum)}$	$R^2_{Y(adj)(cum)}$	Q^2_{cum}	R	p
1	3.237	0.519	0.723	0.688	0.864	1.418×10^{-4}

This study shows that the correlation between observed and predicted $\log(1/EC_{50})$ values is significant ($R = 0.864$, $p < 0.001$) (Table 1). As the cross-validated Q^2_{cum} value of Model (1) is remarkably larger than 0.50, Model (1) is surely stable and has good prediction ability. The model may be used to make predictions for other structure similar PAHs. Based on Model (1), $\log(1/EC_{50})$ values for the other PAHs were predicted (Table 1). It can be seen from Table 1 that the predicted $\log(1/EC_{50})$ values for naphthalene, fluorene, and chrysene are -4.282 nM, -4.546 nM, and -2.629 nM, respectively. The predicted $\log(1/EC_{50})$ values of photocytotoxicity are far lower for naphthalene and fluorene than those for most PAHs, which is comparable to the conclusions of Schirmer et al. (1998a). For chrysene, the predicted EC_{50} values (426 nM) are far greater than its water solubility (13 nM). Schirmer et al. (1998a) concluded that chrysene may not accumulate in cells sufficiently to elicit a photocytotoxic response. Moreover, previous studies showed that direct photolysis quantum yields of chrysene were far lower than those observed for most PAHs (Chen et al. 2000). As the ability to be photocytotoxicity appeared to depend on the proper combination of photochemical and solubility properties (Schirmer et al. 1998a), the predicted result for chrysene may be reasonable. Thus, the predictions may give an initial estimation of the ability of PAHs to be photocytotoxic to a cell line from the rainbow trout gill. Further experimental studies are desired to verify the predictions.

Table 4. The *VIP* values and pseudo-regression coefficients*.

Independent variables	*VIP*	Coeffs (a)	Coeffs (b)
$(E_{LUMO}\text{-}E_{HOMO})^2$	1.209	-2.597×10^{-1}	-3.962×10^{-2}
E_{HOMO}	1.067	2.293×10^{-1}	8.396×10^{-1}
E_{LUMO}	0.998	-2.145×10^{-1}	-8.991×10^{-1}
Mw	0.981	2.109×10^{-1}	5.749×10^{-3}
μ	0.664	-1.427×10^{-1}	-2.498
Constants		-1.891	4.845

* Coeffs (a): Coefficients scaled and centered; Coeffs (b): Coefficients unscaled.

The *VIP* values for the independent variables in Model (1) are listed in Table 4. Table 4 also lists the pseudo-regression coefficients of the independent variables and constants transformed from PLS results. From the positive and negative symbols of the coefficients of the independent variables, one can evaluate the effects of each independent variable on the ability of PAHs to be photocytotoxic to a cell line from the rainbow trout gill, respectively. There are totally five predictor variables included in Model (1). Based on the unscaled pseudo-regression coefficients of the independent variables and constants transformed from PLS results, an analytical QSAR equation can be obtained, as follows:

$$\log(1/EC_{50}) = 4.845 - 3.962\times10^{-2}(E_{LUMO}\text{-}E_{HOMO})^2 + 8.396\times10^{-1}E_{HOMO} - 8.991\times10^{-1}E_{LUMO} + 5.749\times10^{-3}\,Mw - 2.498\mu$$

It can be concluded from this study that the one PLS component is mainly related to the descriptors $(E_{LUMO}\text{-}E_{HOMO})^2$, E_{HOMO}, E_{LUMO}, Mw, and μ. The descriptors $(E_{LUMO}\text{-}E_{HOMO})^2$ and E_{HOMO} are more significant than the other descriptors in governing the $\log(1/EC_{50})$ values of the PAHs. The current results show that increasing $(E_{LUMO}\text{-}E_{HOMO})^2$ values leads decrease of $\log(1/EC_{50})$ values. The $E_{LUMO}\text{-}E_{HOMO}$ gap is related to absolute hardness, defined as half the absolute value of $E_{LUMO}\text{-}E_{HOMO}$ (Pearson 1986), which is regarded as a measure of energy stabilization in chemical systems; chemical structures tend to more stable at larger values of the $E_{LUMO}\text{-}E_{HOMO}$ gap (Faucon et al. 1999; Chen et al. 2001). The present study also shows that PAHs with big $(E_{LUMO}\text{-}E_{HOMO})^2$ values tend to more stable and have small $\log(1/EC_{50})$ values, i.e., the photocytotoxicity of such PAH molecules is low. Increasing μ and E_{LUMO} values of PAHs leads to decrease of the photocytotoxicity. The photocytotoxicity of PAHs increases with the increase of E_{HOMO}. Increasing Mw values of the PAHs leads to increase of $\log(1/EC_{50})$ values. Previous studies (Chen et al. 2001) showed that PAHs with great Mw values tend to unstable to light in the environment, which are consistent with the current phenomena. Schirmer et al. (1998b) found that the ability of PAHs to be directly cytotoxic appeared to be related to their water solubility and to their lipophilicity. Among the PAHs tested, just two- and three-ring PAHs, which have the highest water solubility, were cytotoxic (Schirmer et al. 1998b). Thus it can be concluded that PAHs' cytotoxicity is related to their molecular weight. Compared to the prediction of cytotoxicity, additional parameters, such as $(E_{LUMO}\text{-}E_{HOMO})^2$, E_{HOMO}, E_{LUMO}, and μ which appeared to correlate with the PAHs' photochemical properties (Chen et al. 2001), are important for the prediction of photocytotoxicity.

Acknowledgments. The authors gratefully acknowledge the support from the National Natural Science Foundation (P. R. China, No. 50238020). This project was granted financial support from China Postdoctoral Science Foundation (No. 2003-03).

REFERENCES

Baker J (1986) An algorithm for the location of transition states. J Comput Chem 7:385-395

Boese BL, Lamberson JO, Swartz RC, Ozretich R, Cole F (1998) Photoinduced toxicity of PAHs and alkylated PAHs to a marine infaunal amphipod (*Rhepoxynius abronius*). Arch Environ Contam Toxicol 34:235-240

Chen JW, Peijnenburg WJGM, Quan X, Yang FL (2000) Quantitative structure-property relationships for direct photolysis quantum yields of selected polycyclic aromatic hydrocarbons. Sci Total Environ 246:11-20

Chen JW, Peijnenburg WJGM, Quan X, Chen S, Martens D, Schramm K-W, Kettrup A (2001) Is it possible to develop a QSPR model for direct photolysis half-lives of PAHs under irradiation of sunlight? Environ Pollut 114:137-143

Chen JW, Yang P, Chen S, Quan X, Yuan X, Schramm K–W, Kettrup A (2003)

Quantitative structure-property relationships for vapor pressures of polybrominated diphenyl ethers. SAR QSAR Environ Res 14(2):97-111
Faucon JC, Bureau R, Faisant J, Briens F, Rault S (1999) Prediction of the fish acute toxicity from heterogeneous data coming from notification files. Chemosphere 38:3261-3276
Hu J-Y, Aizawa T (2003) Quantitative structure–activity relationships for estrogen receptor binding affinity of phenolic chemicals. Wat Res 37:1213-1222
Kagan J, Sinnott D, Kagan ED (1987) The toxicity of pyrene in the fish *Pimephales promelas*: synergism by piperonyl butoxide and by ultraviolet light. Chemosphere 16:2291-2298
McClelland HE, Jurs PC (2000) Quantitative structure–property relationships for the prediction of vapor pressures of organic compounds from molecular structures. J Chem Inf Comput Sci 40:967-975
Niu JF, Chen JW, Martens D, Quan X, Yang FL, Kettrup A, Schramm K-W (2003) Photolysis of polycyclic aromatic hydrocarbons adsorbed on spruce (*Picea abies* (L.) Karst.) needles under sunlight irradiation. Environ Pollut 123:39-45
Oris JT, Giesy JP (1986) Photoinduced toxicity of anthracene to juvenile bluegill sunfish (*Lepomis Macrochirus Rafinesque*): photoperiod effects and predictive hazard evaluation. Environ Toxicol Chem 5:761-768
Pearson PG (1986) Absolute electronegativity and hardness correlated with molecular orbital theory. Proc Natl Acad Sci 83:8440-8441
Schirmer K, Chan AGJ, Greenberg BM, Dixon DG, Bols NC (1997) Methodology for demonstrating and measuring the photocytotoxicity of fluoranthene to fish cells in culture. Toxicol Vitro 11:107-119
Schirmer K, Chan AGJ, Greenberg BM, Dixon DG, Bols NC (1998a) Ability of 16 priority PAHs to be photocytotoxic to a cell line from the rainbow trout gill. Toxicol 127:143-155
Schirmer K, Dixon DG, Greenberg BM, Bols NC (1998b) Ability of 16 priority PAHs to be directly cytotoxic to a cell line from the rainbow trout gill. Toxicol 127:129-141
Seward JR, Cronin MTD, Schultz TW (2002) The effect of precision of molecular orbital descriptors on toxicity modeling of selected pyridines. SAS QSAR Environ Res 13:325-340
Stewart JJP (1989a) Optimization of parameters for semiempirical methods Ⅰ. Method. J Comp Chem 10:209-220
Stewart JJP (1989b) Optimization of parameters for semiempirical methods Ⅱ. Applications. J Comp Chem 10:221-264
Veith GD, Mekenyan OG, Ankley GT, Call DJ (1995) A QSAR analysis of substitutent effects on the photoinduced acute toxicity of PAHs. Chemosphere 30:2129-2142
Wold S, Sjöström M, Eriksson L (2001) PLS-regression: a basic tool of chemometrics. Chemom Intell Lab Syst 58:109-130
Yaffe D, Cohen Y (2001) Neural networks based temperature-dependent quantitative structure property relations (QSPRs) for predicting vapor pressure of hydrocarbons. J Chem Inf Comput Sci 41:463-477