Bull. Environ. Contam. Toxicol. (2001) 66:277–282

DOI: 10.1007/s00128-001-0001-6

Bulletin of Environmental Contamination and Toxicology

Quantitative Structure–Property Relationships for Vapor Pressure of PCDD/Fs

J. W. Chen,[1–3,*] X. Quan,[1] Y. Yan,[1] Y. Z. Zhao,[1] F. L. Yang[1]

[1] School of Environmental Science and Technology, Dalian University of Technology, Zhongshan Road 158-129, Dalian 116012, People's Republic of China
[2] State Key Laboratory of Freshwater Ecology and Biotechnology, Wuhan, 430072, People's Republic of China
[3] Institute of Ecological Chemistry, GSF-National Research Center for Environment and Health, D-85764, Neuherberg, Munich, Germany

Received: 15 June 2000/Accepted: 18 December 2000

Polychlorinated dibenzo-*p*-dioxins and dibenzo-*p*-furans (PCDD/Fs) are typical persistent pollutants with high toxicity (Nebert, 1989). Recent studies (Younes 1999; Fossi *et al.* 1999) reveal that most PCDD/Fs are endocrine disrupting chemicals. Vapor pressure is an important parameter that describes the evaporation tendency of an organic pollutant. The development of quantitative structure-property relationship (QSPR) models for vapor pressure is very important for the exposure assessment of PCDD/Fs.

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 QSPR models in which quantum chemical descriptors are used is of great importance. According to the present chemometric theory, as many relevant data as possible should be considered in QSPR studies because this increases the probability of a good characterization of compounds (Kaliszan, 1993). As a consequence of the increase of the number of descriptors, the problem of intercorrelation of independent variables (multicollinearity) will increase. Especially when the number of independent variables is equal to or greater than the number of compounds in the training set, regression analysis (a method that was frequently used in QSPR studies) will not be useful. To overcome these problems, the partial least squares (PLS) method, a widely used chemometric method first developed by Wold *et al.* (1984), will be used in this study.

MATERIALS AND METHODS

In their compilation, Mackay *et al.* (1997) have collected data on solid vapor pressure (P_S) and liquid or subcooled liquid vapor pressure (P_L) for 15 PCDDs and 9 PCDFs. These 24 PCDD/Fs constitute the training set of the study. The logarithms of the vapor pressures are listed in Table 1. In addition, some other PCDD/Fs (mainly those with chlorines substituted in the 2,3,7,8 positions) were selected randomly in the study. The 2,3,7,8- substituted PCDD/Fs are thought to pose a risk to human health due to their toxicity, carcinogenic potency, and potential effects on animal reproductive and immunological systems (Wu, 1999).

Present address: Institute of Ecological Chemistry, GSF-National Research Center for Environment and Health, D-85764, Neuherberg, Munich, Germany
Correspondence to: J. W. Chen

Their P_S and P_L will be predicted based on the models obtained. The predicted values may be useful in exposure assessment of the PCDD/Fs.

Table 1. The PCDD/Fs and their values of $\log P_S$ and $\log P_L$*

No	Compounds	$\log P_S$ (Pa)			$\log P_L$ (Pa)		
		Obs.	*Pred.*	*SE*	*Obs.*	*Pred.*	*SE*
1	Dibenzo-p-dioxin	-1.260	-1.336	±0.178	-0.291	-0.166	±0.155
2	1-CDD	-1.921	-2.427	±0.112	-1.125	-1.344	±0.119
3	2-CDD	-1.770	-1.852	±0.123	-1.137	-1.015	±0.129
4	2,3-D_2CDD	-3.409	-3.437	±0.091	-2.034	-2.220	±0.097
5	2,7-D_2CDD	-3.921	-3.918	±0.132	-2.091	-2.177	±0.098
6	2,8-D_2CDD	-3.854	-3.529	±0.108	-2.607	-2.033	±0.101
7	1,2,4-T_3CDD	-4.000	-4.766	±0.081	-2.971	-3.346	±0.078
8	1,2,3,4-T_4CDD	-5.194	-4.989	±0.085	-3.561	-3.570	±0.076
9	1,2,3,7-T_4CDD	-6.000	-6.188	±0.072	-4.547	-4.259	±0.076
10	1,3,6,8-T_4CDD	-6.155	-6.177	±0.116	-4.236	-3.956	±0.075
11	2,3,7,8-T_4CDD	-6.699	-6.109	±0.116	-3.928	-3.906	±0.075
12	1,2,3,4,7-P_5CDD	-7.056	-6.563	±0.077	-5.374	-4.53	±0.079
13	1,2,3,7,8-P_5CDD		-7.197	±0.088		-4.954	±0.085
14	1,2,3,4,7,8-H_6CDD	-8.292	-8.252	±0.126	-5.839	-5.608	±0.098
15	1,2,3,6,7,8-H_6CDD		-8.311	±0.117		-5.709	±0.100
16	1,2,3,7,8,9-H_6CDD		-7.970	±0.095		-5.601	±0.098
17	1,2,3,4,6,7,8-H_7CDD	-9.125	-9.129	±0.127	-6.752	-6.383	±0.118
18	O_8CDD	-9.959	-10.230	±0.168	-6.021	-7.017	±0.137
19	Dibenzo-p-furan	-0.523	-0.585	±0.181	0.085	-0.304	±0.151
20	2,8-D_2CDF	-3.409	-2.746	±0.124	-1.836	-1.877	±0.105
21	1,2,7,8-T_4CDF		-4.625	±0.169		-3.622	±0.076
22	2,3,7,8-T_4CDF	-5.699	-5.481	±0.087	-3.701	-3.918	±0.075
23	2,3,4,7,8-P_5CDF	-6.456	-6.424	±0.114	-4.764	-4.783	±0.082
24	1,2,3,8,9-P_5CDF		-6.120	±0.162		-4.755	±0.081
25	1,2,3,4,7,8-H_6CDF	-7.495	-7.541	±0.097	-5.511	-5.495	±0.095
26	1,2,3,4,8,9-H_6CDF		-6.742	±0.128		-5.102	±0.087
27	1,2,3,6,7,8-H_6CDF	-7.456	-7.554	±0.119	-5.442	-5.629	±0.099
28	1,2,3,7,8,9-H_6CDF		-7.288	±0.130		-5.496	±0.095
29	2,3,4,6,7,8-H_6CDF		-6.648	±0.116		-4.978	±0.085
30	1,2,3,4,6,7,8-H_7CDF	-8.328	-8.030	±0.115	-6.241	-5.957	±0.107
31	1,2,3,4,7,8,9-H_7CDF	-8.208	-8.519	±0.124	-6.268	-6.326	±0.117
32	O_8CDF	-9.301	-9.328	±0.130	-6.996	-6.924	±0.134

* *Obs.*: Observed values, *Pred.*: Predicted values, *SE*: Standard errors of the predicted values.

PM3 (Stewart 1989a; 1989b) Hamiltonian contained in the quantum chemical computation software MOPAC (Ver. 6.0, *Stewart, J. J. P., 1990, Frank J. Seiler Research Laboratory, U. S. Air Force Academy, Co 80840*) was used to compute quantum chemical descriptors of the PCDD/Fs. The PM3 was selected because it

is a recently developed semi-empirical molecular orbital algorithm and the computational time is much shorter than needed by *ab initio* methods. MOPAC was run with the following keywords: PM3, ESP, POLAR, DIPOLE, PRECISE, NOINTER.

A total of 12 MOPAC derived descriptors reflecting the overall character of the PCDD/F molecules were used in this study. They are molecular weight (*Mw*), average molecular polarizability (α), dipole moment (μ), final heat of formation (*HOF*), total energy (*TE*), electronic energy (*EE*), core-core repulsion energy (*CCR* , *CCR* = *TE* - *EE*), energy of the highest occupied molecular orbital (E_{homo}),

Table 2. Quantum chemical descriptors of the PCDD/Fs

No	*Mw*	α	μ	*HOF*	*TE*	*EE*	E_{homo}	E_{lumo}	Q_C^-	Q_H^+	Q_{Cl}^+
1	184.194	110.156	0	-9.432	-2130.603	-11891.038	-8.645	-0.178	-0.100	0.122	
2	218.639	119.914	0.291	-14.411	-2431.938	-13536.608	-8.751	-0.298	-0.132	0.124	0.102
3	218.639	122.345	0.411	22.603	-2432.001	-13399.609	-8.704	-0.368	-0.125	0.135	0.073
4	253.084	134.107	0.617	-21.122	-2733.349	-15066.905	-8.722	-0.533	-0.136	0.137	0.101
5	253.084	134.935	0.037	-22.179	-2733.395	-14957.399	-8.768	-0.522	-0.124	0.136	0.076
6	253.084	134.897	0.341	-22.202	-2733.396	-14958.712	-8.756	-0.527	-0.125	0.136	0.076
7	287.529	143.019	0.464	-24.491	-3034.614	-17081.506	-8.865	-0.603	-0.150	0.133	0.134
8	321.974	154.804	0.626	-19.989	-3335.921	-19000.732	-8.853	-0.709	-0.141	0.125	0.139
9	321.974	157.106	0.250	-31.974	-3336.026	-18660.313	-8.840	-0.745	-0.151	0.139	0.137
10	321.974	155.775	0.041	-31.554	-3336.040	-18690.131	-8.935	-0.722	-0.121	0.139	0.115
11	321.974	159.537	0.023	-32.573	-3336.084	-18441.711	-8.800	-0.785	-0.136	0.138	0.106
12	356.419	167.900	0.306	-35.039	-3637.310	-20736.372	-8.898	-0.822	-0.141	0.127	0.141
13	356.419	169.555	0.167	-36.395	-3637.369	-20492.895	-8.855	-0.854	-0.151	0.139	0.139
14	390.865	180.536	0.032	-40.182	-3938.653	-22625.924	-8.905	-0.922	-0.140	0.139	0.142
15	390.865	179.648	0.031	-40.182	-3938.653	-22609.132	-8.912	-0.920	-0.151	0.140	0.139
16	390.865	179.086	0.314	-40.160	-3938.652	-22650.455	-8.905	-0.916	-0.152	0.140	0.142
17	425.310	190.133	0.153	-43.930	-4239.935	-24847.674	-8.960	-0.981	-0.151	0.140	0.146
18	459.755	200.645	0.037	-47.668	-4541.216	-27151.547	-9.009	-1.037	-0.140		0.146
19	168.195	106.918	0.326	25.491	-1837.164	-10106.423	-9.018	-0.477	-0.128	0.122	
20	237.085	130.157	0.187	12.422	-2439.969	-13066.484	-9.099	-0.785	-0.150	0.125	0.068
21	305.975	151.973	0.402	29.168	-3042.640	-16650.267	-9.051	-1.041	-0.162	0.137	0.115
22	305.975	157.256	0.044	3.178	-3042.669	-16417.580	-9.032	-1.079	-0.161	0.138	0.105
23	340.420	168.124	0.244	-2.062	-3343.956	-18340.863	-9.068	-1.169	-0.161	0.139	0.137
24	340.420	160.504	0.472	-0.288	-3343.879	-18742.164	-9.097	-1.138	-0.179	0.140	0.126
25	374.865	178.161	0.059	-6.571	-3645.271	-20451.212	-9.020	-1.273	-0.163	0.139	0.140
26	374.865	172.240	0.363	-4.084	-3645.163	-20756.808	-9.048	-1.235	-0.168	0.129	0.139
27	374.865	177.322	0.140	-6.598	-3645.272	-20439.361	-9.098	-1.263	-0.174	0.140	0.137
28	374.865	175.059	0.271	-4.537	-3645.214	-20603.679	-9.058	-1.264	-0.179	0.140	0.130
29	374.865	178.805	0.408	-5.858	-3645.240	-20352.829	-9.119	-1.254	-0.150	0.127	0.140
30	409.310	188.909	0.244	-10.347	-3946.554	-22532.548	-9.074	-1.351	-0.163	0.133	0.144
31	409.310	186.811	0.105	-9.242	-3946.506	-22679.458	-9.046	-1.351	-0.179	0.141	0.142
32	443.755	198.324	0.087	-12.969	-4247.787	-24845.576	-9.089	-1.429	-0.168		0.144

energy of the lowest unoccupied molecular orbital (E_{lumo}), the largest negative atomic charge on a carbon atom (Q_C^-), the most positive net atomic charges on a hydrogen atom (Q_H^+), and the most positive net atomic charges on a chlorine atom (Q_{Cl}^+). 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 kilocalories, and units of energy, charge, dipole and polarizability are electron volts (eV), atomic charge units (a.c.u) and atomic units (a.u.) respectively.

Simca (Simca-S Version 6.0, *Umetri AB & 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 fitted values (R), and the significance level (p).

RESULTS AND DISCUSSION

PLS analysis for the 24 PCDD/Fs in the training set, with logP_S and logP_L as dependent variables and the 12 quantum chemical descriptors as independent variables, resulted in QSPR models (1) and (2) respectively. The results of the 2 PLS models are listed 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. For example, it can be concluded from Table 3 that 2 PLS principal components were selected in model (1), and the 2 PLS principal components explained 74.7% of the variance of the independent variables, and 98.4% of the variance of the dependent variable.

Table 3. Model fitting results

Models	k	$R^2_{X(adj)(cum)}$	$R^2_{Y(adj)(cum)}$	Q^2_{cum}	R	p
(1) logP_S	2	0.747	0.984	0.980	0.993	7.496×10^{-22}
(2) logP_L	1	0.627	0.970	0.963	0.986	1.824×10^{-18}

As shown by Figure 1 and Figure 2, for the PCDD/Fs under study, the correlation between observed and predicted logP_S or logP_L values is very significant. As the cross-validated Q^2_{cum} values of model (1) and model (2) are remarkably larger than 0.50, the two models are surely stable and have a good prediction ability. The models may be used to make predictions for other PCDD/Fs. Based on the models, logP_S and logP_L values for other PCDD/Fs under study were predicted (Table 1).

VIP (Variable Importance in the Projection) is a parameter that shows the importance of a variable in a model. Terms with a large value of *VIP*, larger than

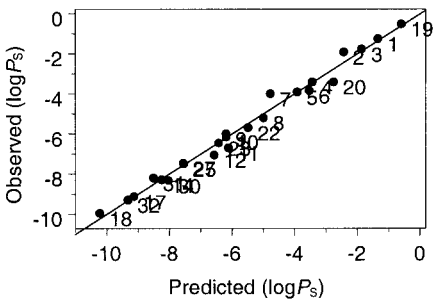

Figure 1. Plot of observed logP_S values versus those predicted by model (1)

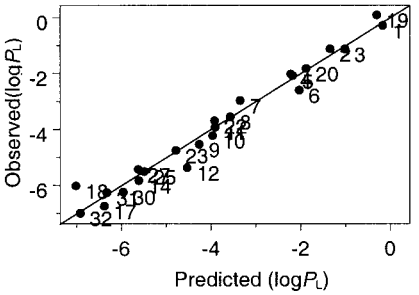

Figure 2. Plot of observed logP_L values versus those predicted by model (2)

1, are the most relevant for explaining the dependent variable. The *VIP* values for the independent variables in model (1) and model (2) 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 dependent variables. Based on the unscaled coefficients and constants, QSPR equations like those obtained from multiple regression analysis can be obtained.

Table 4. *VIP*s (Variable Importance in the Projection) and pseudo-regression coefficients (Unscaled)

	Model (1)			Model (2)	
Variables	*VIP*	*Coefficients*	Variables	*VIP*	*Coefficients*
α	1.205	-1.235×10^{-2}	α	1.205	-9.027×10^{-3}
Mw	1.202	-4.136×10^{-3}	*Mw*	1.201	-3.033×10^{-3}
TE	1.197	4.959×10^{-4}	*TE*	1.190	3.492×10^{-4}
EE	1.180	7.779×10^{-5}	*EE*	1.171	5.520×10^{-5}
CCR	1.176	-9.212×10^{-5}	*CCR*	1.166	-6.547×10^{-5}
Q_{Cl}^{+}	1.074	-8.025	Q_{Cl}^{+}	1.151	-8.833
E_{lumo}	1.038	6.863×10^{-1}	E_{lumo}	1.067	6.243×10^{-1}
Q_H^{+}	0.851	-3.659×10	Q_C^{-}	0.859	9.456
Q_C^{-}	0.805	4.702	Q_H^{+}	0.815	-2.612×10
E_{homo}	0.719	8.642×10^{-1}	E_{homo}	0.730	1.077
HOF	0.708	1.613×10^{-2}	*HOF*	0.625	6.779×10^{-3}
μ	0.523	1.291	μ	0.414	4.712×10^{-1}
Constant		1.684×10	*Constant*		1.769×10

From the data in Table 4, one may conclude the following: (I) The descriptors α, *Mw*, *TE*, *EE*, *CCR*, Q_{Cl}^{+} and E_{lumo} are more significant than the other descriptors in governing the logP_S and logP_L values of the PCDD/Fs. (II) Increasing α, *Mw*, and *CCR* values of the PCDD/Fs leads to decreasing logP_S and logP_L values, while increasing *TE* and *EE* values of the PCDD/Fs leads to increasing logP_S and logP_L values. This is because these descriptors are inter-correlated. α correlates with *Mw* and *CCR* positively, while it correlates with *TE* and *EE* negatively. Thus, the more

chlorine atoms present in the parent structures, the bigger the molecules, the greater the α values, and thus the smaller the $\log P_S$ and $\log P_L$ values. This is reasonable since intermolecular dispersive forces are in direct proportion to α^2, and the intermolecular dispersive forces do not favor volatilization. (III) Increasing Q_{Cl}^+ and E_{lumo} values of the PCDD/Fs leads to a decrease in $\log P_S$ and $\log P_L$ values. This is reasonable since Q_{Cl}^+ and E_{lumo} characterize the molecular abilities to donate and/or accept electrons in intermolecular interactions and PCDD/F molecules with great Q_{Cl}^+ and E_{lumo} values tend to have great intermolecular interactions that do not favor volatilization.

Acknowledgments: The study was supported by the Teaching and Research Award Program for Outstanding Young Teachers in Higher Education Institutions of MOE (TRAPOYT), P. R. China, and the State Key Laboratory of Freshwater Ecology and Biotechnology (2000FB12), P. R. China. The research results were attained with the assistance of the Alexander von Humboldt (AvH) Foundation and we thank the AvH foundation. We thank Mr. Christopher Alan Noser at Carnegie Mellon University for reviewing our manuscripts.

REFERENCES

Fossi MC, Casini S, Marsili L (1999) Nondestructive biomarkers of exposure to endocrine disrupting chemicals in endangered species of wildlife. Chemosphere 39:1273-1285

Kaliszan R (1993) Quantitative structure retention relationships applied to reversed-phase high-performance liquid chromatography. J Chromatogr 656A:417-435

Mackay D, Shiu WY, Ma KC (1992) Illustrated handbook of physical-chemical properties and environmental fate for organic chemicals. Vol. I-V. Lewis publishers, Boca Raton, Ann Arbor, London.

Nebert DWC (1989) The Ah locus: genetic differences in toxicity cancer mutations and birth defects. CRC Crit Rev Toxicol 20:153-174

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

Wold S, Wold H, Dunn W J III (1984) Report UMINF-83, Department of Chemistry, University of Umeå, Sweden

Wu WZ (1999) Environmental Behavior and Ecotoxicological Impact of Persistent Organic Pollutants (POP) in Wildlife, with Special Emphasis on the Aquatic Ecosystem, Munich: Herbert Utz Verlag, Wissenschaft

Younes M (1999) Specific issues in health risk assessment of endocrine disrupting chemicals and international activities. Chemosphere 39:1253-1257