# QUANTITATIVE PREDICTIVE MODELS FOR OCTANOL-AIR PARTITION COEFFICIENTS OF PERSISTENT ORGANIC POLLUTANTS AT DIFFERENT ENVIRONMENTAL TEMPERATURES

Jingwen Chen[1,2], Tom Harner[3], K.-W. Schramm[2], A. Kettrup[2,4]

[1] School of Environmental Science and Technology, Dalian University of Technology, Zhongshan Road 158-129, Dalian 116012, P. R. China
[2] Institute of Ecological Chemistry, GSF-National Research Center for Environment and Health, D-85764, Neuherberg, Germany
[3] Environment Canada, Meteorological Service of Canada, 4905 Dufferin Street, Downsview Ontario M3H 5T4, Canada
[4] Lehrstuhl für Ökologische Chemie und Umweltanalytik, Technische Universität München, 85350 Freising-Weihenstephan, Germany

## Introduction

The octanol-air partition coefficient ($K_{OA}$) is a key descriptor of chemicals partitioning between the atmosphere and organic phases[1]. Recently, $K_{OA}$ based approaches have been successfully employed to model surface-air partitioning of persistent organic pollutants to aerosols[2-7], soil[8-9], vegetation[10-11], and even indoor carpet[12]. Because of the large enthalpy change involved in octanol to air transfer, $K_{OA}$ has strong temperature ($T$) dependence, which is very important for assessing the global transportation of persistent organic pollutants (POPs). The experimental determination of $K_{OA}$ needs special equipment and samples, is expensive and time and labor consuming. It is therefore desirable to develop a reliable predictive technique for determining $K_{OA}$.

A chemical's physical chemical properties and environmental partitioning is related to its molecular structure. In this study molecular structural descriptors are used and develop predictive models for polychlorinated naphthalenes (PCNs), chlorobenzenes (CBs), polychlorinated biphenyls (PCBs) and polychlorinated dibenzo-*p*-dioxin/dibenzofurans (PCDD/Fs).

## Methods and Materials

To date, $K_{OA}$ values at different environmental temperatures have been determined by the generator column method[13] for selected CBs[13], PCBs[13,14], PCNs[15] and PCDDFs[1]. It was based on these experimental values that the quantitative predictive models were to be developed.

CS ChemOffice (Version 6.0) and the PM3 Hamiltonian[16] 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 molecular structural descriptors. The molecular structural descriptors used in the model development are listed in Table 1.

Partial least squares (PLS) analysis[17] was used in developing the models. Simca (Simca-S Version 6.0, *Umetri AB & Erisoft AB*) software was used to perform the PLS analysis. The criterion used to determine the model dimensionality is cross validation. With cross validation, 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

**Table 1.** List of the molecular structural descriptors

| Descriptors | Meanings |
|---|---|
| $Mw$ | Molecular weight |
| a | Average molecular polarizability |
| $\mu$ | Dipole moment |
| *ÄH*$_{\mathrm{f}}$ | Standard heat of formation |
| $TE$ | Total energy |
| $EE$ | Electronic energy |
| $CCR$ | Core-core repulsion energy |
| $E_{\mathrm{HOMO}}$ | The energy of the highest occupied molecular orbital |
| $E_{\mathrm{LUMO}}$ | The energy of the lowest unoccupied molecular orbital |
| $q_{\mathrm{C}}^{-}$ | The largest negative net atomic charge on a carbon atom |
| $q_{\mathrm{H}}^{+}$ | The most positive net atomic charges on a hydrogen atom |
| $q_{\mathrm{Cl}}^{+}$ | The most positive net atomic charges on a chlorine atom |
| $q^{+}$ | The most positive net atomic charge on hydrogen or chlorine atoms in PCDD molecules |
| $CAA$ | Connolly accessible area |
| $CMA$ | Connolly molecular area |
| $CSEV$ | Connolly solvent-excluded volume |
| $Ov$ | Ovality |

components ($k$), $Q^2_{\mathrm{cum}}$, the correlation coefficient between observed and fitted values ($R$), and the significance level ($p$). Besides these statistics, a standard error ($SE$) was adopted to characterize the prediction precision of models. $SE$ was defined like that in multiple regression analysis, i.e.,

$$SE = \sqrt{\frac{\sum_{i=1}^{n}[\log K_{\mathrm{OA}}(\mathrm{observed}))_i - \log K_{\mathrm{OA}}(\mathrm{predicted})_i]^2}{n-k-1}}$$

where $n$ stands for the number of observations in the training set.

The chlorinated compounds under study were divided into three groups for the study: (1) CBs, PCNs, p,p'-DDT; (2) PCBs; and (3) PCDD/Fs. For each group, some compounds with reported $K_{\mathrm{OA}}$ values were selected to constitute a validation set. For the validation set of group (1), $K_{\mathrm{OA}}$ values were determined using the generator column method[13,15]. For the validation set of group (2), $K_{\mathrm{OA}}$ values were determined semi-empirically using a multi-column method[18]. The $K_{\mathrm{OA}}$ values at 25 ºC extrapolated by simple regression analysis[1] (Validation set I) and determined semi-empirically by retention time indices using equation 2 of Harner *et al.*[1] (Validation set II) constitute the validation set of group (3).

## Results and Discussion

By PLS analysis, model (1), (2) and (3) were obtained respectively for the three groups of polychlorinated compounds.

Model (1) (For CBs, PCNs and p,p'-DDT):

$\log K_{\mathrm{OA}} = 2.543\times10 - 4.843\times10^{-4}TE - 8.339\times10^{-5}EE + 9.916\times10^{-5}CCR + 4.203\times10^{-3}Mw + 1.060\times10^{-2}\,\alpha$ -

$6.914\times10^{-1}E_{LUMO} + 1.117q_{Cl}^{+} - 1.386q_{C}^{-} + 9.126\times10^{-1}E_{HOMO} - 7.715\times10^{-11}T^{4} - 3.009\times10^{-8}T^{3} - 1.312\times10^{-5}T^{2}$
$- 7.585\times10^{-3}T - 2.177\ln T + 6.211\times10^{2}/T + 1.270\times10^{-2}DH_{f} - 8.274q_{H}^{+} + 2.644\times10^{-1}\mu$
$n = 99, k = 3, Q^{2}_{cum} = 0.981, R = 0.992, p = 8.777\times10^{-89}, SE = 0.152$

Model (2) (For PCBs):
$\log K_{OA} = -5.507\times10^{2}E_{LUMO}/T + 3.599\alpha/T + 1.052CAA/T + 1.473CMA/T - 4.693\times10^{-2}TE/T + 4.048\times10^{-1}Mw/T + 1.849\times10^{2}q_{Cl}^{+}/T + 1.768\times10^{-1}CSEV/T - 1.30710^{-3}EE/T + 6.44410^{-5}CCR/T - 3.544\times10^{3}q_{C}^{-}/T + 1.013\times10^{3}Ov/T - 1.892\times10^{3}q_{H}^{+}/T - 6.828DH_{f}/T + 6.892E_{HOMO}/T + 5.257\times10^{2}/T - 4.393$
$n = 87, k = 4, Q^{2}_{cum} = 0.976, R = 0.990, p = 1.044\times10^{-73}, SE = 0.187$

Model (3) (For PCDD/Fs):
$\log K_{OA} = 1.650\times10^{-2}\alpha + 5.863\times10^{-3}Mw - 3.720\times10q^{+} - 6.139\times10^{-4}TE - 1.476E_{LUMO} + 2.196\times10^{3}/T - 7.320\ln T + 4.164\times10$
$n = 41, k = 3, Q^{2}_{cum} = 0.983, R = 0.993, p = 3.641\times10^{-38}, SE = 0.132$

It can be found that in Model (1) and (3), the molecular structural descriptors were included directly. Whereas in Model (2), the temperature adjusted molecular structural descriptors were found to be appropriate.

As the cross-validated $Q^{2}_{cum}$ values of models (1)~(3) is remarkably larger than 0.50, these three models are surely stable and have a good prediction ability. As shown by Figures 1, 2 and 3, for the compounds in training set and validation set, the predicted values are consistent with the corresponding determined values. It has been validated that the three models can be used for predictions of $K_{OA}$ for other structurally similar compounds within the same class.

Based on the *VIP* (Variable Importance in the Projection) values and the pseudo-regression coefficients included in models (1)~(3), the following conclusions can be obtained. (I) For the POPs contained in group (1) and (3), the main factors governing $\log K_{OA}$ are, in order of decreasing importance are: molecular size, environmental temperatures and molecular ability of accepting/ donating electrons. For PCBs: environmental temperatures, $E_{LUMO,}$ molecular size, and the net atomic charges on chlorine, hydrogen and carbon atoms of PCB molecules, are major molecular structural factors governing $\log K_{OA}$. (II) For CBs and PCNs: dipole-dipole, dipole-induced dipole, and induced dipole-induced dipole intermolecular interactions may be the main factors governing the magnitude of

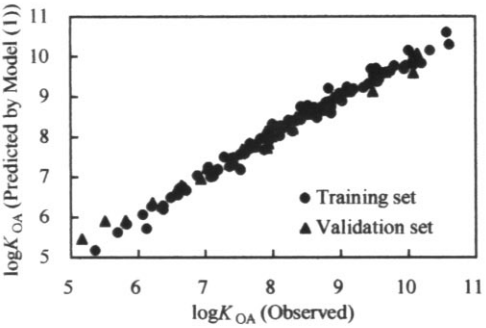

**Figure 1.** Plot of observed $\log K_{OA}$ values versus those predicted by model (1)

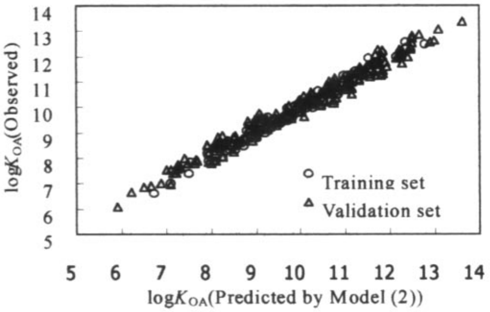

**Figure 2.** Plot of observed $\log K_{OA}$ values versus those predicted by model (2)

temperature dependence of log$K_{OA}$. For PCBs, $q_C^-$ and $\alpha$ are major structural factors governing temperature dependence of log$K_{OA}$.

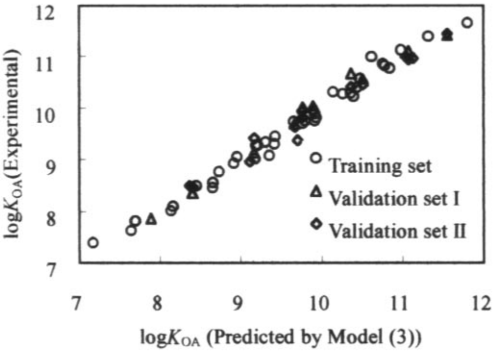

**Figure 3.** Plot of observed log$K_{OA}$ values versus those predicted by model (3)

**Ackowledgments**

The study is supported by the Teaching and Research Award Program for Outstanding Young Teachers in Higher Education Institutions of MOE (TRAPOYT), P. R. China. The research results were attained with the assistance of the Alexander von Humboldt Foundation.

**References**

1. Harner T., Green N. J. L. and Jones K. C. (2000) Environ Sci Technol. 34, 3109.
2. Finizio A., Mackay D., Bidleman T. and Harner T. (1997) Atmos Environ. 31, 2289.
3. Harner T. and Bidleman T. F. (1998) Environ Sci Technol. 32, 1494.
4. Kaupp H. and McLachlan M. S. (1999) Chemosphere, 38, 3411.
5. Lee R. G. M. and Jones K. C. (1999) Environ Sci Technol. 33, 3596.
6. Lohmann R., Lee R. G. M., Green N. J. L. and Jones K. C. (2000) Atmos Environ. 34, 2529.
7. Falconer R. L. and Harner T. (2000) Atmos Environ. 34, 4043.
8. Hippelein M. and McLachlan M. S. (1998) Environ Sci Technol. 32, 310.
9. Cousins I. T., Beck A. J. and Jones K. C. (1999) Sci Total Environ. 228, 5.
10. Thomas G., Sweetman A. J., Ockenden W. A., Mackay D. and Jones K. C. (1998) Environ Sci Technol. 32, 936.
11. Weiss P. (2000) Environ Sci Technol. 34, 1707.
12. Won D., Corsi R. L. and Kynes M. (2000) Environ Sci Technol. 34, 4193.
13. Harner T. and Mackay D. (1995) Environ Sci Technol. 29, 1599.
14. Harner T. and Bidleman T. F. (1996) J Chem Eng Data, 41, 895.
15. Harner T. and Bidleman T. F. (1998) J Chem Eng Data, 43, 40.
16. Stewart J. J. P. (1989) J Comput Chem. 10, 209.
17. Wold S., Wold H. and Dunn WJIII. (1984) Report UMINF-83, Department of Chemistry, University of Umeå, Sweden
18. Zhang X. M., Schramm K–W., Henkelman B., Klimm G., Kaune A., Kettrup A, and Lu P. Z. (1999) Anal Chem. 71, 3834.