PREDICTING RATE CONSTANTS FOR GAS PHASE REACTIONS OF HYDROXYL RADICAL WITH PAHS, PCDD/FS, PCBS AND PBDES AT DIFFERENT TEMPERATURES

Wang Y N[1], Chen J W[1*], Li X H[1], Qiao X L[1]

[1]Key Laboratory of Industrial Ecology and Environmental Engineering (MOE), Department of Environmental Science and Technology, Dalian University of Technology, Linggong Road 2, Dalian 116024, China

Abstract
The gas phase reaction with hydroxyl radical ($\cdot$OH) is the most important removal process in the day time for organic pollutants in the atmosphere, thus the $\cdot$OH reaction rate constants (k_{OH}) are of great importance to characterize the fate of organic pollutants in the troposphere. In this study, a new quantitative structure-activity relationship (QSAR) model was developed for predicting k_{OH} of polycyclic aromatic hydrocarbons (PAHs), polychlorinated dibenzo-p-dioxins/dibenzofurans (PCDD/Fs), polychlorinated biphenyls (PCBs), and polybrominated diphenyl ethers (PBDEs) over the temperature range of 294~388 K, applying 8 parameters and partial least squares (PLS) regression. The QSAR development followed OECD guidelines, with specific attention to validation, applicability domain and mechanistic interpretation. For the optimal model, the leave-many-out cross-validated $Q^2_{CUM} = 0.872$, $R^2 = 0.887$, $RMSE = 0.174$ log units, and the external validated $Q^2_{EXT} = 0.828$, indicating good robustness and predictive capability. The applicability domain of the model was analyzed by Williams plot, mostly covering halogenated aromatic compounds. The key factors governing $\log k_{OH}$ are molecular ability to donate electrons, as described by the energy of the highest occupied molecular orbital (E_{HOMO}), the number of halogen atoms in a molecule (n_X), and molecular weight (M_W) and the environmental temperature ($1/T$ and $\log T$).

Introduction
Polycyclic aromatic hydrocarbons (PAHs), polychlorinated dibenzo-p-dioxins/dibenzofurans (PCDD/Fs), polychlorinated biphenyls (PCBs) and polybrominated diphenyl ethers (PBDEs) are organic pollutants that are widely distributed geographically, persistent and mobile in the environment. These pollutants also tend to be bio-accumulating and toxic, and therefore they pose risks to biota and humans. Most PAHs are mutagenic and carcinogenic, while PCDD/Fs, PCBs as well as PBDEs are included in the ''Stockholm Convention of Persistent Organic Pollutants (POPs)''. These pollutants are mainly released into the atmosphere from various anthropogenic sources, such as combustion sources.

It is well known that the gas phase reaction of organic pollutants with $\cdot$OH in the atmosphere is the primary process for their degradation and transformation in the daytime.[1] $\cdot$OH reaction rate constants (k_{OH}) are of great importance to assessing the behavior and fate of organic pollutants in the troposphere. To date, there have been some experiments performed to determine the k_{OH} values of PAHs, PCDD/Fs, PCBs and PBDEs. However, the experimental methods are time-consuming, laborious, costly and equipment dependent, thus it is crucial to develop predictive methods for k_{OH}.

The present study is to develop a universal quantitative structure-activity relationship (QSAR) model for k_{OH} prediction of PAHs, PCDD/Fs, PCBs and PBDEs at different temperatures, following the OECD principles. According to the OECD consensus, to facilitate the consideration of a QSAR model for regulatory purposes, it should be associated with the follow information: a defined endpoint; an unambiguous algorithm; a defined domain of applicability; appropriate measures of goodness-of-fit,

---

* Correspondence to: Jingwen Chen, e-mail: jwchen@dlut.edu.cn, Tel/Fax: +86 - 411 - 8470 6269.

robustness and predictive power; and a mechanistic interpretation, if possible.[2] Therefore, we collected 413 experimental k_{OH} values from literatures, employing partial least squares (PLS) regression for model development. The key factors governing k_{OH} were analyzed and the temperature dependence of $\log k_{OH}$ was also discussed.

Materials and methods

A total of 413 k_{OH} values (in units of cm^3 $molecule^{-1}$ s^{-1}, 294~388 K and 1 atm) for 21 PAHs, 6 PCDD/Fs, 18 PCBs and 7 PBDEs were collected, which were experimentally determined by different researchers and laboratories.[3-6] The average k_{OH} values for some a PAH, PCDD/F, PCB or PBDE determined by different laboratories and researchers at a certain temperature, were calculated and then transformed to logarithmic units ($\log k_{OH}$) with the standard deviation (SD) varying between 0 and 0.17, averaging 0.03. Thus the database is composed of 190 $\log k_{OH}$ values ranging from -12.07 to -9.74 log unit, with mean = -11.19 log unit. The database was then randomly split into a training set (155 objects) for model development, and a validation set (35 objects) for external validation. The straightforward random selection technique was employed for creating the two subsets. A full list of the summarized experimental k_{OH} can be obtained from the corresponding author by e-mail.

Since quantum chemical descriptors have clear physical chemical definitions, which can make QSAR models more transparent, a total of 9 quantum chemical descriptors were selected for preliminary model development.[7] Molecular weight (Mw) was also added to describe the molecular size. These parameters were computed by MOPAC 2000 in the CS Chem3D Ultra (Version 6.0) using the PM3 Hamiltonian method. Moreover, our previous study proved that molecular atom distributions and compactness (Ds), the number of halogen atoms (n_X) in a molecule and the information of functional groups adjacent to the reaction sites (BELm2) were the most influential descriptors for gas-phase k_{OH}, thus these 3 descriptors were also included and calculated by DRAGON (Version 5.5).[7] Besides, temperature (T, 1/T, logT) were added to characterize the temperature dependence of $\log k_{OH}$.

To develop QSAR models for $\log k_{OH}$, PLS regression was performed by using Simca (Version 6.0, Umetri AB & Erisoft AB). The cross-validated Q^2_{CUM} was employed to indicate the robustness and internal predictivity of the PLS model, which is the cumulative variance of the dependent variable that can be explained by the extracted PLS components. Besides, the correlated coefficient square (R^2), the minimum root mean squared error (RMSE) were employed to express the robustness of developed models. To obtain the optimum PLS model, the PLS analysis was repeated to eliminate irrelevant variables using Q^2_{CUM} and variable importance in the projection (VIP) as indicators. The optimal model was then evaluated by external validation, using Q^2_{EXT} as the quantification of the predictive capability. The Williams plot, the plot of the standardized residuals versus the leverage, was exploited to visualize the applicability domain (AD).[7]

Results and Discussion

The analytical form of the optimal PLS model is:

$$\log k_{OH} = 1.86 + 1.31 E_{HOMO} - 1.70 \times 10^{-1} n_X + 4.98 \times 10^{-4} Mw - 6.06/T + 5.60 \times 10^{-1} \log T + 3.19 \times 10^{-4} \Delta H_f + 2.60 \times 10^{-1} E_{LUMO} - 1.39 \, BELm2 \tag{1}$$

$$n = 155, A = 3, k = 8, Q^2_{CUM} = 0.827, R^2 = 0.840, RMSE = 0.207, p < 0.001.$$

where n stands for number of objects in the training set, A is the number of PLS components, k is the number of predictor variables, and p is the significant level. Q^2_{CUM} of the optimal model is 0.827, indicating the goodness-of-fit and robustness. For the validation set of 35 objects, $Q^2_{EXT} = 0.794$, RMSE = 0.224. Figure 1 plots the predicted versus experimental $\log k_{OH}$ values for the training set and the validation set, which indicates the good predictive capability of the optimal model.

Williams plot (Figure 2) shows that only one chemical (1-ethylnaphthalene) in the training set with

leverage > warning leverage ($h* = 0.174$) and the standardized residuals $<3\sigma$, which fit the model well and make the model more precise. The analysis of the applicability domain highlighted no presence of outliers.
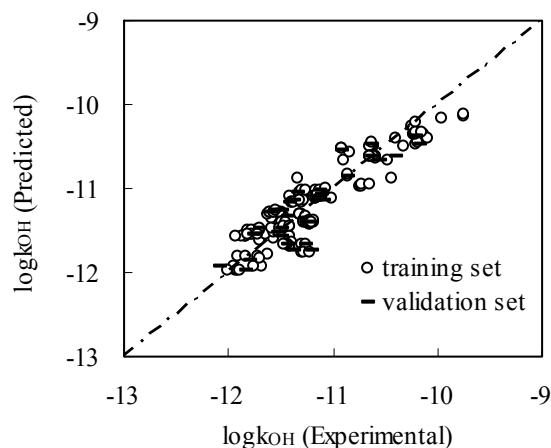

Figure 1. Plots of predicted versus experimental $\log k_{OH}$ values for the training and validation set.
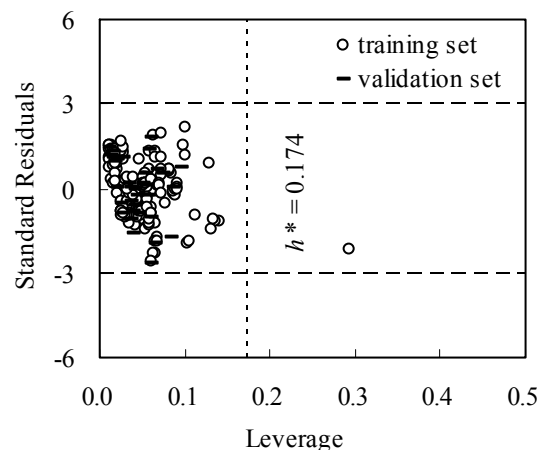
Figure 2. Plots of standardized residuals versus leverages. Dash lines represent $\pm$ 3 standardized residuals, dotted line represents warning leverage ($h* = 0.174$).

Model (1) extracts 3 PLS components that relevant to 8 parameters (Table 1). The *VIP* value (especially *VIP* > 1) shows the importance of each parameter in model (1). From the PLS weights ($W*[1]$, $W*[2]$, $W*[3]$), it can be seen how much a single predictor variable contributes in each PLS component to the modeling of $\log k_{OH}$. The $\pm$ signs of parameter coefficients and the PLS weights inform about the correlation between response variable and predictor variables. Thus the key factors governing k_{OH} can be analyzed by *VIP*, coefficients and PLS weights.

Table 1. Parameters, their *VIP* values of 8 parameters, regression coefficients, and PLS weights in Model (1).

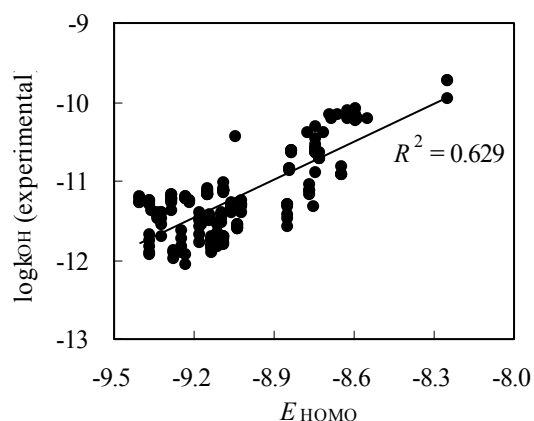|	*VIP*	coefficients	$W*[1]$	$W*[2]$	$W*[3]$
E_{HOMO}	**1.59**	1.31	**0.59**	0.66	0.47
n_X	**1.46**	-0.17	**-0.56**	-0.23	-0.32
Mw	**1.35**	3.04×10^{-3}	**-0.49**	0.16	0.63
$1/T$	0.66	-6.06	0.18	**-0.47**	0.03
$\log T$	0.66	0.56	-0.17	**0.50**	0.03
ΔH_f	0.53	1.24×10^{-3}	0.19	-0.15	-0.01
E_{LUMO}	0.50	0.26	0.02	-0.10	-0.46
$BELm2$	0.33	-1.39	-0.07	-0.23	0.48

Figure 3. Plots of experimental $\log k_{OH}$ v.s. E_{HOMO}.

The first PLS component accounts for 50.01% variance of $\log k_{OH}$, which is loaded primarily on the descriptors E_{HOMO}, n_X and Mw. E_{HOMO} (VIP_{max}) characterizes the susceptibility of a molecule toward attack

by electrophile. Its coefficient and the positive $W*[1]$ value (Table 1) indicate that as a general trend $\log k_{OH}$ increases with increasing E_{HOMO} (Figure 3). For ·OH is an electrophilic reagent, molecules with higher abilities of donating electrons (E_{HOMO}) tend to be more reactive towards ·OH. n_X characterizes the number of halogen atoms. Taking PCBs (4-chlorobiphenyl, 4,4'-dichlorobiphenyl, 2,4,4'-trichlorobiphenyl, 2,2',4,4'-tetrachlorobiphenyl) for examples, the molecules with more halogen atoms tend to have lower $\log k_{OH}$ values (-11.31, -11.60, -11.65, -11.94, respectively; temperature: 343 K). Mw encodes the information of the molecular size.

The second PLS component accounts for 22.76% variance of $\log k_{OH}$, which is loaded primarily on the descriptors $1/T$, $\log T$ with low VIP values, implying the weak influence of the temperature to $\log k_{OH}$, which agrees with the previous experimental results of Brubaker and Hites as well as the models analysis of Yan et al.[3-5] Besides, the results indicated that temperature dependence of $\log k_{OH}$ related with $1/T$ and $\log T$. It is consistent with the temperature-dependent rate expression, which can be given as the Arrhenius expression $k_{OH} = Ae^{-B/T}$ or as three-parameter expression $k_{OH} = AT^n e^{-B/T}$. When rate constants have been measured over a range of temperatures, Arrhenius plots of $\log k_{OH}$ versus $1/T$ often exhibit curvature and hence the temperature dependence expression are recommended in terms of three-parameter expression $k_{OH} = AT^n e^{-B/T}$.[8] For rate constants employed in present study over a temperature range of 294~388 K, it makes sound that the temperature dependence of $\log k_{OH}$ can be expressed in terms of $1/T$ and $\log T$.

References
1. Aschmann S.M., Arey J., Atkinson R., Simonich S.L. *Environ Sci Technol* 2001; 35: 3595.
2. OECD, Organisation for Economic Co-Operation and Development, Paris, France, 2007.
3. Yan C.L., Chen J.W., Huang L.P., Ding G.H., Huang X.Y. *Chemosphere* 2005; 61: 1523.
4. Brubaker W.W., Hites R.A. *Environ Sci Technol* 1997; 31: 1805.
5. Anderson P.N., Hites R.A. *Environ Sci Technol* 1996; 30: 301.
6. Raff J.D., Hites R.A. *J Phys Chem* 2006; 110: 10783.
7. Wang Y.N., Chen J.W., Li X.H., Wang B., Cai X.Y., Huang L.P. *Atmos Environ* 2009; 43: 1131.
8. Atkinson R. Kinetics of the gas-phase reactions of OH radicals with alkanes and cycloalkanes. *Atmos Chem Phys Discuss* 2003; 3: 4183.