THE QSPR STUDY OF THE SAMPLING RATES OF SEMIPERMEABLE MEMBRANE DEVICE FOR POLYCYCLIC AROMATIC HYDROCARBONS

Zhu X H[1], Ding G H[2], Levy W[3], Wang W[1], Mu J[1], Dong X W[1], Schramm K-W [3,4]

[1]School of Environmental and Chemical Engineering, Dalian Jiaotong University, Dalian 116028, China

[2]College of Environmental Science and Engineering, Dalian Maritime University, Linghai Road 1, Dalian 116026, China

[3]Helmholtz Zentrum München - German Research Center for Environmental Health, Institute of Ecological Chemistry, Ingolstädter Landstr. 1, D-85764 Neuherberg, Munich, Germany

[4]TUM, Wissenschaftszentrum Weihenstephan fuer Ernaehrung und Landnutzung, Department fuer Biowissenschaftliche Grundlagen, Weihenstephaner Steig 23, 85350 Freising, Germany

Abstract

QSPR models were developed for air sampling rates (R_{air}) of semipermeable membrane devices (SPMDs) for PAHs, using partial least square (PLS) regression. Quantum chemical descriptors computed by semi-empirical PM6 method were used as predictor variables. The cumulative variance of the dependent variable explained by the PLS components and determined by cross-validation (Q^2_{cum}), for the optimal models, is 0.799, indicating that the model has good predictive ability and robustness, and could be used to estimate R_{air} values of the above mentioned compounds. The main factors governing R_{air} values of PAHs are intermolecular interactions with weak electron-transfer and the energy required for cave-forming in dissolution of PAHs into triolein of SPMDs.

Introduction

Semi-permeable membrane devices (SPMDs) is one of the popular passive samplers for sampling persistent organic pollutants (POPs) in air[1-2]. The application SPMDs has been summarized in many articles [1-4]. By means of SPMDs sampling, only the amount of POPs sequestrated in each SPMD can be determined. However, a main community concern is the current concentrations of POPs in air. How can be determined the current air concentrations from the amount of sequestered POPs in SPMDs? When SPMDs sampling is in linear uptake stage, the atmospheric concentration of POPs, C_{air} (pg m^{-3}), can be calculated from the air sampling rate of SPMDs, R_{air} (m^3 day^{-1} SPMD^{-1}), the concentrations of POPs sequestered by SPMDs (N_{SPMD}, pg SPMD^{-1}) and the exposure time of SPMDs (t,days) using the following equation[5]:

$$C_{air} = N_{SPMD}/(R_{air}\ t) \tag{1}$$

It has been found that R_{air} were influenced by the molecular weight, the substitution patterns of chlorine in polychlorinated biphenyls(PCBs), and K_{oa} of POPs[2] [6-8]. It is well known that quantitative structure–property relationship (QSPR) method provides a convenient tool to predict physicochemical properties of chemicals only from molecular structural information, and it may also provide insight into main factors that influence physicochemical properties of chemicals[9-10]. To our knowledge there is no QSPR study about R_{air} to date. The aim of this paper is to develop QSPR model for R_{air} of polycyclic aromatic hydrocarbons (PAHs), and analyze the key factors influencing the R_{air}. It is expected that an ideal model can be found which can be used to predict R_{air} only from the structure characteristics of PAHs. Then in the future, only a few amount of R_{air} of PAHs should be measured, while the R_{air} of other PAHs could be predicted by the QSPR model, and consequently the C_{air} could be derived using equation (1).

Materials and Methods

Data Source. PAHs data were taken from the literature [11]. The atmospheric concentration of each PAH used to calculate R_{air} was the total air concentration (C_v+C_p, where C_v is the concentration of the compound in the vapor phase; C_p is the concentration of the compound in particle phase). The sampling rates of SPMDs calculated according to the total air concentration for PAHs are listed in Table 1.

Descriptors and statistical analysis. Recently, semiempirical, more complete parameter optimization PM6 method[12] has been implemented in MOPAC 2009 package[13], giving satisfactory estimates of molecular properties. In this paper, molecular structures of the studied compounds were drawn with CS ChemDraw Ultra 6.0. Then geometry optimizations and quantum-mechanical calculations were performed at the semi-empirical PM6 level[12] with the MOPAC 2009 package[13]. A total 17 molecular descriptors that reflect the overall characteristics of the chemicals were obtained from the MOPAC output files. The molecular structural descriptors are listed in Table 2.

Table 1. SPMD sampling rates of PAHs

PAH	R_{air} ($m^3 d^{-1} SPMD^{-1}$)
Fluorene (Flo)	1.6
phenanthrene (Phe)	4.2
anthracene (Ant)	3.7
fluoranthene (Flu)	4.3
pyrene (Pyr)	1.8
benz[a]anthracene (B[a]A)	1.5
chrysene (Chr)	3.1
benzo[e]pyrene (B[e]P)	1.2
benzo[a]pyrene (B[a]P)	0.5
indeno[1,2,3-c,d]pyrene (I[c,d]P)	1.7
benzo[g,h,i]perylene (B[g,h,i]P)	0.7

Partial least squares (PLS) regression was adopted for constructing QSPR model, because this method can analyze data with strongly collinear, noisy and numerous predictor variables[14]. In this study, the PLS analysis was carried out by Simca-S (Version 6.0, Umetri AB & Erisoft AB). Simca-S employs "cross validation" to determine the number of PLS components (A). Q^2_{cum} is a main statistic parameter of a PLS model, which denotes the cumulative variance of the dependent variable explained by the extracted PLS components, and is a good measure of the predictive power and robustness of the model. When Q^2_{cum} of a model is larger than 0.5, the model can be considered to be predictive and robust. In order to obtain the optimal QSPR model, a variable selection procedure described in the literature [9] was adopted here.

Results and Discussion

PLS analyses were performed with all the descriptors. After variable selection, the optimal PLS model obtained was:

$$R_{air} = -2.187 \times 10^{-1} - 1.714 \times 10^{-3} CCR - 3.485 E_{HOMO} + 1.795 \times 10^{-1} q_C^- \qquad (2)$$
$$n = 11, A = 2, R^2_{X(adj)(cum)} = 0.614, R^2_{Y(adj)(cum)} = 0.841, Q^2_{cum} = 0.799, r = 0.934$$

Where n is the number of data points, A is the number of PLS components, $R^2_{X(adj)(cum)}$ and $R^2_{Y(adj)(cum)}$ stand for the cumulative variance of all the predictor variables and dependent variable, respectively, explained by the extracted components, and r is the correlation coefficient between observed and fitted values.

2

Table 2. Theoretical molecular structural descriptors

No.	Descriptors	Descriptions
1	ΔH_{f}	Standard heat of formation (kJ)
2	TE	Total energy (eV)
3	EE	Electronic energy (eV)
4	CCR	Core–core repulsion energy (eV)
5	IP	the vertical ionization potential (eV)
6	DE	the dielectric energy (eV)
7	A_{COSMO}	COSMO area (Å^2)
8	V_{COSMO}	COSMO volume(Å^3)
9	Mw	molecular mass (atomic mass units)
10	ME	Mulliken's electronegativity (eV)
11	AH	Parr and Pople's absolute hardness (eV)
12	E_{HOMO}	The energy of the highest occupied molecular orbital (eV)
13	E_{LUMO}	The energy of the lowest unoccupied molecular orbital (eV)
14	q_{C}^{-}	The most negative net atomic charges on a carbon atom (atomic charge unit)
15	q_{H}^{+}	The most positive net atomic charges on a hydrogen atom (atomic charge unit)
16	μ	The total dipole moment(Debye)
17	α	average molecular polarizability (atomic units)

Fig. 1 is the plot of predicted vs. observed R_{air} according to equation (2), from which it can be seen that the predicted R_{air} values are consistent with the observed values. Q^2_{cum} of the optimal models is 0.799, which is higher than 0.5, indicating that the model has good predictive ability and robustness, and could be used to estimate R_{air} values for this kind of compounds. In view of difficulties in the measurements of R_{air}, the predicted data could be regarded as an approximation of air sampling rates. Thus C_{air} for the sampling site can be obtained by calculating with N_{SPMD} and R_{air}, where there are no possibilities for active sampling measurements.

y = 0.8724x + 0.2819
r = 0.9341
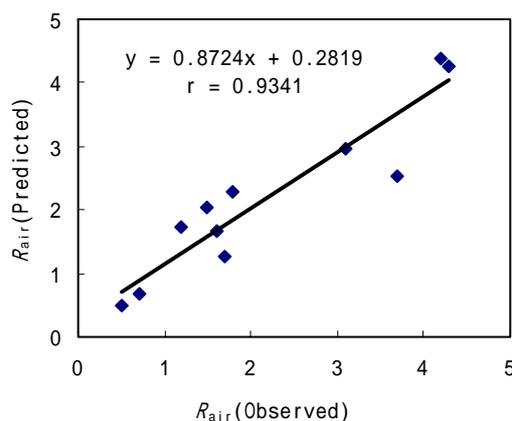
Fig.1. Plot of observed and predicted R_{air} values.

Variable importance in the projection (*VIP*) of a predictor variable indicates its influence on response variable,

3

which is listed in Table 3. The values of weights of predictor variables (w^*) and weights of response variable (c) of the PLS model are also presented in Table 3. The weights can represent how much a single variable contributes in each PLS component to the modeling of R_{air}. Predictor variables which are more important for the nth PLS component, have higher $w^*[n]$ absolute values, and response variables well modeled by the nth PLS component have large $c[n]$ absolute values. From the values of w^* in Table 3, it can be seen that the first PLS component is mainly relevant to E_{HOMO} and CCR, and the second PLS component q_C^-. CCR is significantly related to Mw and α, and may represent non-specific intermolecular interactions or the energy required for cave-forming in dissolution of PAHs into triolein of SPMDs. From its coefficient, it can be seen that CCR is negatively correlated with R_{air}, so it may mainly describe the energy required for cave-forming, which is a key factor for R_{air} of PAHs. E_{HOMO} and q_C^- characterize molecular abilities to accept or donate electrons in intermolecular interactions, and therefore are related to intermolecular interactions with weak electron-transfer. The stronger the intermolecular interactions are, the higher the air sampling rate is.

Table 3. VIP values and PLS weights in model

	VIP	w*c[1]	w*c[2]
E_{HOMO}	1.104	-0.701	-0.099
CCR	1.042	-0.662	-0.095
q_C^-	0.833	0.266	1.075
R_{air}		0.717	0.395

References

1. Esteve-Turrillas F.A., Pastor A., Yusà V., Guardia M. de la. *Trends in Analytical Chemistry* 2007; 26:703.

2. Huckins J.N., Petty J.D., Booij K. Monitors of Organic Chemicals in the Environment – Semipermeable Membrane Device. 2006;Springer Science,New York, USA.

3. Zhu X., Pfister G., Henkelmann B., Kotalik J., Fiedler S., Schramm K.-W. *Chemosphere* 2007a; 68: 1623.

4. Zhu X., Pfister G., Henkelmann B., Kotalik J., Bernhoft S., Fiedler S., Schramm K.-W. *Environ. Pollut.* 2008; 156: 461.

5. Ockenden W.A., Sweetman A.J., Prest H.F., Steinnes E., Jones K. *Environ. Sci. Technol.* 1998b; 32: 2795.

6. Liu G.Q., Zhang G., Li J., Li X.D., Peng X.Z., Qi S.H. *Atmos. Environ.* 2006; 40: 3134.

7. Lohmann R., Corrigan B.P., Howsam M., Jones K.C., Ockenden W.A. *Environ. Sci. Technol.* 2001; 35: 2576.

8. Ockenden W.A., Prest H.F., Thomas G.O., Sweetman A., Jones K.C. *Environ. Sci. Technol.* 1998a; 32: 1538.

9. Ding G., Chen J., Qiao X., Huang L., Lin J., Chen X. *Chemosphere* 2006; 62: 1057.

10. Zhu X.H., Wang W., Schramm K.-W., Niu W. *Chromatographia* 2007b; 65:719.

11. Barkow M.E., Huckins J.N., Mu J.F. *Atmos. Environ.* 2004,38:5938.

12. Stewart J.J.P. *J. Mol. Modeling* 2007; 13: 1173.

13. Stewart J.J.P. 2008. http://OpenMOPAC.net [01-12-2008].

14. Wold S., Sjöström M., Eriksson L. *Chemom. Intell. Lab. Syst.* 2001; 58: 109.

4