Prediction of octanol-air partition coefficients for PCBs at different ambient temperatures based on the solvation free energy and the dimer ratio

Wanran Li [a], Guanghui Ding [a, *], Hong Gao [b], Yuting Zhuang [a], Xiaoyue Gu [a], Willie J.G.M. Peijnenburg [c, d]

[a] College of Environmental Science and Engineering, Dalian Maritime University, Linghai Road 1, Dalian, 116026, PR China
[b] Department of Physics, Dalian Maritime University, Linghai Road 1, Dalian, 116026, PR China
[c] Center for Safety of Substances and Products, National Institute of Public Health and the Environment, P.O. Box 1, Bilthoven, the Netherlands
[d] Institute of Environmental Sciences (CML), Leiden University, Leiden, 2300, the Netherlands

HIGHLIGHTS

- A predictive model for log K_{OA} of PCBs was developed based on $\triangle G_{OA}$.
- The optimal combination of theoretical method and basis-set was HF/MIDI!6D.
- The model was improved after taking the dimer formation into account.
- Log K_{OA} values of PCBs at different ambient temperatures were predicted.

ARTICLE INFO

ABSTRACT

Temperature-dependent octanol-air partition coefficients (K_{OA}) are of great importance in assessing the environmental behavior and fate of persistent organic pollutants including polychlorinated biphenyls (PCBs). Due to the tremendous amounts of time, effort and cost needed for the experimental determination of K_{OA}, it is desirable to develop a rapid and precise predictive method to estimate K_{OA} just based on molecular structure. In the present study, a predictive model for log K_{OA} of PCBs at ambient temperatures was developed based on the thermodynamic relationship between K_{OA} and the solvation free energy from air to octanol (ΔG_{OA}). For the calculation of ΔG_{OA} of PCBs, the optimal combination of theoretical method and basis-set was identified to be HF/MIDI!6D for both geometry optimization and energy calculation. Dimer formation could affect the partition behavior and promote the apparent K_{OA} values of PCBs. After taking the effect of dimer formation into account, the goodness-of-fit, predictive ability, and robustness of the predictive model were significantly improved. Apparent log K_{OA} values of PCBs at different ambient temperatures ranging from 283.15 to 303.15 K were predicted. Compared with other reported models, the model developed in the present study had not only comparable goodness-of-fit and predictive ability, but also a universal application domain and the relative independency of experimental data. Therefore, the solvation free energy method could be a promising method for the prediction of K_{OA}.

© 2019 Elsevier Ltd. All rights reserved.

1. Introduction

PCBs are of great concern due to their environmental persistence, bioaccumulation, long-range transport potential and toxicity.

Among many parameters affecting their environmental behavior and fate, the octanol-air partition coefficient (K_{OA}) is a key physicochemical property characterizing the volatility and the partitioning behavior between air and environmental organic phases. However, due to the tremendous amounts of time, effort and cost needed for the experimental determination of K_{OA}, it is not well possible to directly measure the K_{OA} values of the whole set of 209 PCB congeners. In addition, due to pragmatic problems of

availability of pure standards, purity of chemicals, suitable analytical methods, different measurement methods, and experimental difficulties like enhanced sorption of chemicals to the wall of the measuring device, it is in itself challenging if not impossible to experimentally determine all K_{OA} values of PCBs. Therefore, it is desirable to develop a rapid and precise prediction method to estimate this essential physicochemical property.

Numerous prediction methods for K_{OA} have been reported, such as the K_{OW}-HLC method (Meylan and Howard, 2005), the solvation parameters model (Abraham and Al-Hussaini, 2005), quantitative structure-activity relationships (QSARs) (Chen et al., 2002), and the thermodynamic method based on the solvation free energy (Fu et al., 2016). The first three methods greatly depend on corresponding experimental data. According to the thermodynamic relationship (Eq. (1)),

$$\log K_{OA} = \frac{-\Delta G_{OA}}{2.303RT} \tag{1}$$

where R (8.314 J mol^{-1}·K^{-1}) is the gas constant, and T is the absolute temperature (K), a time-saving method could be developed to estimate log K_{OA} values directly from the solvation free energy of a chemical from air to octanol (ΔG_{OA}). As ΔG_{OA} could be accurately calculated based on quantum chemistry nowadays, the thermodynamic method based on the solvation free energy is less dependent on experimental data than other methods. In addition, log K_{OA} values at different ambient temperatures could be estimated due to the slight change of ΔG_{OA} in a narrow temperature range.

With the development of quantum chemistry, many continuum solvation models have been developed and widely used to calculate the solvation free energy, such as the polarized continuum model (PCM) (Miertus et al., 1981), the real solvents conductor-like shielding model (COSMO-RS) (Kholod et al., 2011; Klamt, 2005, 2011), COSMO-SAC models (Hsieh et al., 2010; Lin and Sandler, 2002; Phillips et al., 2011), and SMx models (Cramer and Truhlar, 2008; Gupta et al., 2012; Marenich et al., 2013; Thompson et al., 2004). Among them, Solvation Model Density (SMD) model, a novel and effective continuum solvation model, has been developed by Marenich et al. (2009) to calculate the solvation free energy. The performance of the SMD model varied with the combination of different theoretical methods and basis-sets for molecular geometry optimization and energy calculation. Different theoretical methods, such as HF (Levine, 1983), B3LYP (Stephens et al., 1994), M05-2X (Zhao et al., 2006), and M06-2X (Zhao and Truhlar, 2008a), and basis-sets, such as 6-31G(d) (Petersson et al., 1988), 6-31G(d,p) (Petersson and Al-Laham, 1991), 6-31+G(d,p) (Clark et al., 1983), cc-pVTZ (Kendall et al., 1992) and MIDI!6D (Easton et al., 1996) have been used in the SMD model to calculate the solvation free energy. However, the optimal combination of a theoretical method and a basis-set for the calculation of ΔG_{OA} of PCBs is not clear.

In addition, it has been pointed out that hydrophobic organic pollutants could be present in the form of monomers, dimers and polymers, instead of just individual molecules, in environmental matrices (Guo et al., 2015; Janda et al., 1975; Li and Chen, 2014; Shen et al., 2018). During the experimental measurement of K_{OA}, concentrations of PCBs are relatively high, and thus PCB molecules have a high probability to encounter each other to form dimers and polymers. Therefore, the experimental log K_{OA} values of PCBs may be the combined values of a mixture of monomers, dimers and polymers, rather than the values of individual molecules. As dimers and polymers are more hydrophobic than individual molecules, the actually measured log K_{OA} value are probably higher than the theoretical value of individual molecules. However, little attention

has been given to the influence of dimer and polymer formation on the theoretical prediction of physicochemical properties, such as log K_{OA}.

Among dimers and various polymers, hydrophobic organic pollutants are likely to exist mainly in the form of dimers because of their very low concentrations in the environment (Wild et al., 2008). Nowadays, there are a number of experimental and theoretical evidences for the formation of dimers of benzene and PAHs (Arunan and Gutowsky, 1993; Chakraborty and Lim, 1993; Doxtader et al., 1986; Gonzalez and Lim, 2000; Janda et al., 1975; Lee et al., 2019; Miliordos et al., 2014; Piacenza and Grimme, 2005; White et al., 1998). It has been shown that the driving force for the formation of dimers of benzene and PAHs, comes from the π-π interaction (Hunter and Sanders, 1990; Hwang et al., 2015; Miliordos et al., 2014; Sherrill, 2013). Chemicals could form dimers as long as there are aromatic π-systems in the molecular structure. Thus, PCBs could also form dimers in the environment although there is no direct experimental evidence. Therefore, for the theoretical prediction of log K_{OA} of PCBs, taking monomers and dimers into account without polymers should be desirable, and could simplify the calculations.

Therefore, in the present study, the optimal combination of a theoretical method and a basis-set for calculating ΔG_{OA} of PCBs was explored, and a thermodynamic model based on the solvation free energy was developed to estimate log K_{OA} values of PCBs at different temperatures. Furthermore, the influence of dimer formation on the prediction of log K_{OA} was analyzed and amended to obtain precise estimates of log K_{OA} values of PCBs.

2. Materials and methods

2.1. Data set

The experimental log K_{OA} values of 36 PCBs at 298.15 K were taken from the publications of Harner and Mackay (1995), Harner and Bidleman (1996), Kömp and McLachlan (1997), and Wania et al. (2002). The experimental values ranged from 6.80 to 10.51 and were listed in Table S1.

2.2. Prediction of log K_{OA} based on the solvation free energy

Molecular structures of each PCB congener and dimer were drawn by CS ChemDraw Ultra (Version 14.0, Cambridge Scientific Computing, Inc.). Calculation of ΔG_{OA} based on the SMD model involved three steps, geometry optimization, frequency calculation, and energy calculation, which were all conducted with Gaussian 09-E01 software (Frisch et al., 2009). The frequency calculation was conducted to guarantee the optimized geometry was surely the minimum energy point at the potential energy surface. The singe-point energies in the gas phase and the octanol phase were calculated respectively by using the SMD model based on different combinations of theoretical methods and basis-sets. Finally, the log K_{OA} values were estimated from ΔG_{OA} with Eq. (1).

2.3. Theoretical methods and basis-sets

Four theoretical calculation methods and five basis-sets were used for the geometry optimization and energy calculation in the present study. The four theoretical calculation methods were HF, B3LYP, M05-2X, and M06-2X, respectively. HF is an ab initio method, while B3LYP, M05-2X, and M06-2X are based on density functional theory. Five basis-sets were adopted as follows: 6-31G(d), 6-31G(d,p), 6-31+G(d,p), cc-pVTZ and MIDI!6D. These theoretical calculation methods and basis-sets have been

recommended to be used for the parameterization of the SMD model (Marenich et al., 2009). In the present study, a total of 100 combinations of the theoretical calculation method and the basis-set were obtained (Table S2), and tested for geometry optimization and energy calculation. Subsequently the optimal combination was selected to calculate ΔG_{OA}.

2.4. Estimation of the ratio of monomeric and dimeric PCBs

In environmental matrices, PCBs are likely to exist mainly in the form of monomer and dimer. It was therefore assumed that the experimental log K_{OA} value of a PCB was the sum of the values of the corresponding monomeric and dimeric PCB. The ratios of monomeric and dimeric PCBs were then estimated based on the experimental log K_{OA} values and the predicted log K_{OA} values of monomeric and dimeric PCBs. Subsequently, QSAR models were developed to predict the ratios of monomer to dimer of individual PCBs without experimental log K_{OA} values. For the QSAR model, 11 molecular structural descriptors were selected, as they had been previously used for prediction of log K_{OA} (Chen et al., 2003, 2004). These descriptors included core-core repulsion (*CCR*, eV), dipole moment (μ, atomic unit), heat of formation (nH_f, kJ/mol), ionization potential (*Ip*, eV), molecular weight (*Mw*), total energy (*TE*, eV), polarizability (α, atomic unit), Connolly accessible area (*CAA*, Å^2), Connolly molecular area (*CMA*, Å^2), Connolly solvent excluded volume (*CSEV*, Å^3) and Ovality. The PM7 method (Stewart, J.J.P., 2013) in MOPAC 2016 (Stewart Computational Chemistry, Colorado Springs, Co, USA) was employed to calculate these molecular structure descriptors. The QSAR models were developed by using partial least-squares (PLS) regression in Simca-S (Version 13.0, Umetri AB & Erisoft AB) software following the modeling method of Ding et al. (2006).

2.5. Statistical analysis methods

The predictive properties of the models were characterized by the coefficient of determination (R^2), root mean square error (*RMSE*) and leave-one-out cross-validation statistic (Q^2_{CV}). The *F*-test and the *t*-test were used to test the statistical significance of the models and the explanatory variables, respectively. In PLS analysis, the variable importance in the project (*VIP*) reflected the relative importance of an independent variable in a prediction model. A variable with *VIP* value close to or greater than 1 can be considered important in a given model. Variables with *VIP* scores far less than 1 are unimportant and are more likely to be excluded from the model.

Abnormal data points, including outliers, high leverage points and influential points, were identified based on the studentized residual (r_i), leverage value (p_{ii}) and Cook's distance (C_i), respectively, as calculated by means of the following equations:

$$r_i = \frac{e_i}{\hat{\sigma}_{(i)}\sqrt{1 - p_{ii}}} \tag{2}$$

$$p_{ii} = \frac{1}{n} + \frac{(x_i - \overline{x})^2}{\sum (x_i - \overline{x})^2} \tag{3}$$

$$C_i = \frac{\sum_{j=1}^{n}(\hat{y}_j - \hat{y}_{j(i)})^2}{\hat{\sigma}^2(n+1)} \tag{4}$$

where e_i is the residual error of i; $\hat{\sigma}_{(i)}$ is the estimate of standard deviation σ, calculated by regression fitting with the remaining $n-1$ observations after deleting the observation of i; x_i is the observed

value of i; $\overline{x}$ is the average of the observed values; $\hat{y}_j$ is the predicted value of j; $\hat{y}_{j(i)}$ is the predicted value of j after removing the predicted value of i; and n is the sample size.

3. Results and discussion

3.1. Calculation of ΔG_{OA} of PCBs

For 36 PCBs, for which the experimental log K_{OA} values at 298.15 K were available, the experimental values of ΔG_{OA} were calculated with Eq. (1). ΔG_{OA} values of these 36 PCBs were also predicted by the SMD model with different combinations of the theoretical calculation method and the basis-set (also called as calculation levels). Based on the experimental and predicted ΔG_{OA} values, 100 simple linear regression models were developed, and R^2 values were obtained for different combinations of calculation levels. These R^2 values reflected the goodness-of-fit of the models and were presented in Table S3. With the R^2 values, the best simple linear regression model and the corresponding optimal combination of calculation levels could be chosen. It can be seen from Table S3 that the highest R^2 was equal to 0.843, as obtained by a HF/MIDI!6D calculation level for both the geometry optimization and single-point energy calculation.

The simple linear regression model between the experimental and predicted ΔG_{OA} values with the highest R^2 was termed Model 1. The equivariance and normal distribution of residuals were verified by the standardized residual scatter plot, standard residual histogram and normal probability plot (Fig. S1). No outliers were identified for Model 1 based on the studentized residual, as shown in Fig. S2A. However, PCB-3 (4-PCB) and PCB-126 (3,3′,4,4′,5-PCB) were identified as high leverage points from Fig. S2B. It is known that a high leverage point may not necessarily be an influential point. Therefore, Cook's distances were subsequently calculated to identify influential points. Based on the Cook's distance, PCB-3 and PCB-61 (2,3,4,5-PCB) were recognized as influential points for Model 1, while PCB-126 was not recognized as an influential point (Fig. S2C). It is also known that an influential data point could greatly affect the slope of a regression line. Therefore, PCB-3 and PCB-61 were removed for further analyses, but not PCB-126.

After PCB-3 and PCB-61 were removed, an improved model was obtained, named as Model 2. From Fig. S3A, it can be seen that the predicted ΔG_{OA} of Model 2 agreed well with the experimental values. The equivariance and normal distribution of residuals were also verified by the standardized residual scatter plot, standard residual histogram and normal probability plot (Fig. S3). No outliers were identified for Model 2 (Fig. S4A). PCB-126 was identified to be a high leverage point again, while it was not recognized as an influential point (Fig. S4). From the Cook's distance, PCB-29 (2,4,5-PCB) and PCB-5 (2,3-PCB) were found to be influential points for Model 2 (Fig. S4C). Therefore, PCB-29 and PCB-5, not PCB-126, were subsequently removed for further analyses.

After PCB-29 and PCB-5 were removed, Model 3 was built with the goodness-of-fit improved. The standardized residual scatter plot (Fig. S5B), the standard residual histogram (Fig. S5C) and the normal probability plot (Fig.S5D) of Model 3 confirmed the equivariance and normal distribution of residuals. No outliers or high leverage points were identified for Model 3 (Fig. S6A and Fig. S6B). While PCB-8 (2,4′-PCB) and PCB-171 (2,2′,3,3′,4,4′,6-PCB) were identified to be influential points for Model 3 (Fig. S6C). Therefore, PCB-8 and PCB-171 were removed for further analyses.

Model 4 was obtained after PCB-8 and PCB-171 were removed. The values of R^2 and F (1, 28) of Model 4 were 0.885 and 214.051 ($P < 0.01$), respectively, indicating the significant goodness-of-fit of Model 4. Fig. 1A showed that the predicted ΔG_{OA} values were
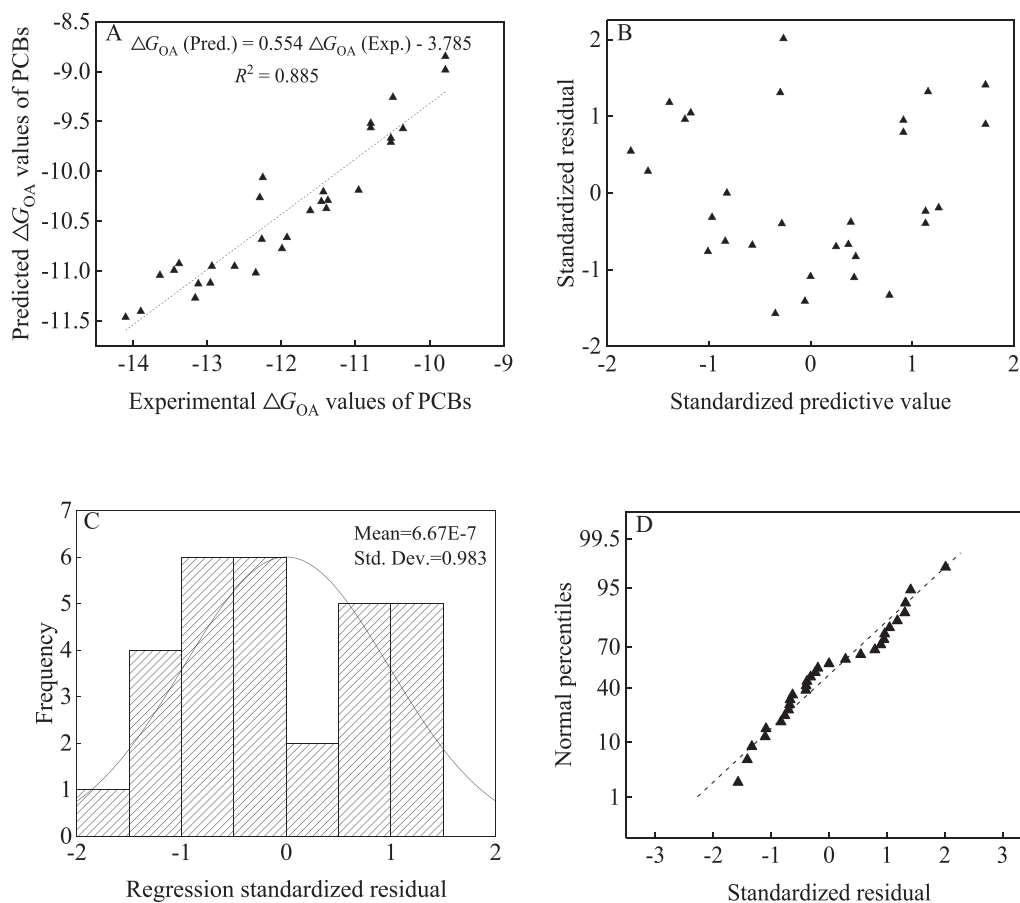
Fig. 1. Experimental and predicted ΔG_{OA} values of PCBs (A), the standardized residual scatter plot (B), standard residual histogram (C) and normal probability plot (D) of Model 4.

consistent with the experimental values. The result of the Durbin-Watson test showed that no self-correlation existed and the residuals were independent. As illustrated in the standardized residual scatter plot (Fig. 1B), the data points were uniformly distributed without any special distribution shape. Both the standard residual histogram (Fig. 1C) and the normal probability plot (Fig. 1D) indicated that the standard residuals approximately followed a normal distribution.

As shown in Fig. 2A, no outliers were identified for Model 4. However, PCB-4 (2,2′-PCB), PCB-10 (2,6-PCB) and PCB-126 were recognized to be high leverage points, as their leverage values were higher than the warning value (Fig. 2B). From Fig. 2C, it could be seen that PCB-10 was an influential point, without PCB-4 and PCB-126. Re-building the model with PCB-10 removed, led to the decrease of R^2 and Q^2_{CV} from 0.885 to 0.874 and from 0.877 to 0.867, respectively. Both the changes of R^2 and Q^2_{CV} indicated that the goodness-of-fit and the predictive ability of the model declined after removing PCB-10. Therefore, PCB-10 might not be an abnormal point, and was not removed from Model 4.

The statistics of these models were shown in Table 1. It can be seen that Model 4 had the highest R^2 and Q^2_{CV} values, and the lowest *RMSE*, indicating the high goodness-of-fit, robustness and predictive ability of Model 4. Thus, the optimal combination of the theoretical method and basis-set for geometry optimization and energy calculation were identified again with abnormal points removed. After abnormal points were removed, R^2 values and the corresponding combination of calculation levels were presented in Table S4. It can be seen that the optimal combination of the theoretical method and basis-set for geometry optimization and energy

calculation was in both cases HF/MIDI!6D. Therefore, HF/MIDI!6D was chosen for the calculation of ΔG_{OA} of PCBs.

3.2. Comparison of theoretical methods

Keeping the basis-set the same, the effects of calculation methods on the prediction of ΔG_{OA} were evaluated. As shown in Fig. 3, the order of the calculation methods was HF > B3LYP > M06-2X > M05-2X in most cases. However, when 6-31G(d,p) was used for the geometry optimization, results of the calculation method of B3LYP were better than those of HF.

The HF method is based on three approximations (non-relativistic approximation, adiabatic approximation and orbital approximation), and neglects the electron correlation energy (Levine, 1983). B3LYP describes the properties of a system based on density functional theory rather than wave function, and takes the electron correlation energy into account (Stephens et al., 1994). In view of these considerations, B3LYP should possess a higher accuracy than HF. However, the energy calculated by B3LYP could be underestimated due to excessive consideration of electron delocalization (Tirado-Rives and Jorgensen, 2008). Therefore, the predictive accuracy of HF was generally superior to the accuracy of B3LYP in this study. M05-2X and M06-2X are highly parametrized, empirical exchange correlation functionals, which have been shown to better describe non-covalent interactions than density functionals (Hohenstein et al., 2008). These two methods were claimed to also capture "medium-range" electron correlation, but were found to ignore the "long-range" electron correlation (Hohenstein et al., 2008). Therefore, for calculating the solvation
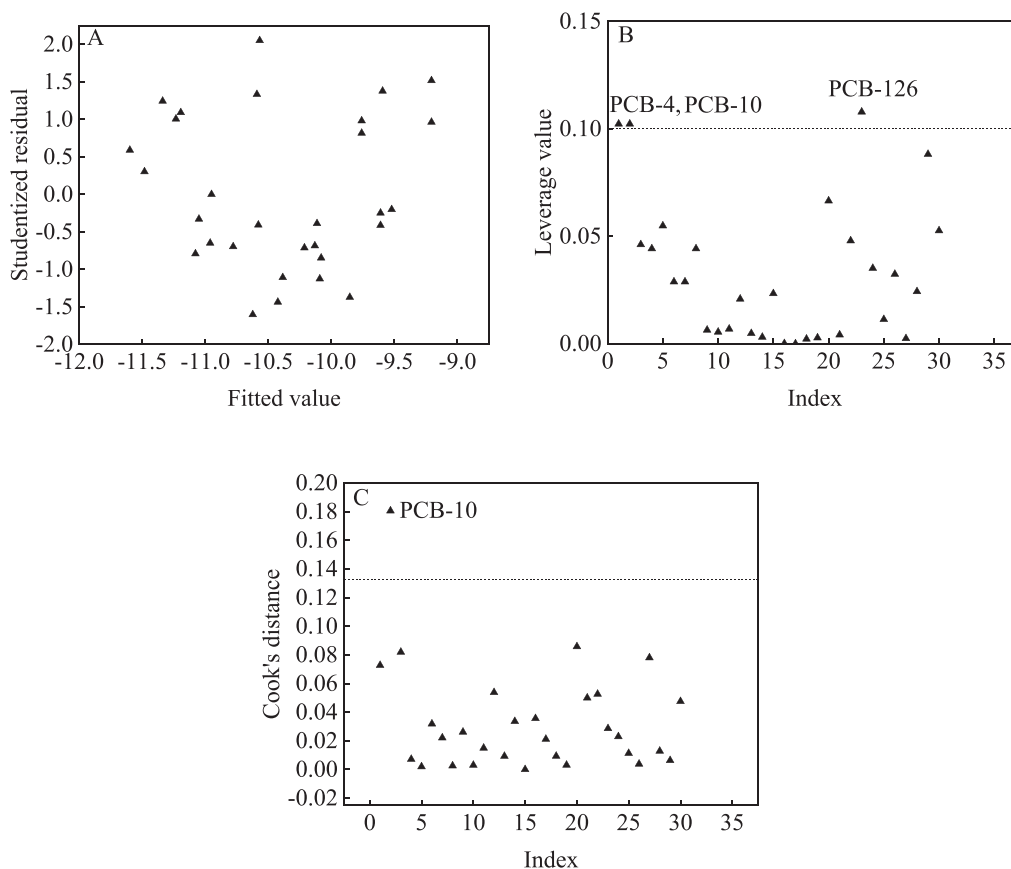
Fig. 2. The studentized residual (A), leverage value (B) and Cook's distance (C) of PCBs in Model 4.

Table 1
Statistics of some simple linear regression models developed in this study.

Model	n	R^2	RMSE	Q^2_{CV}
Model 1	36	0.843	0.346	0.836
Model 2	34	0.866	0.299	0.858
Model 3	32	0.881	0.261	0.873
Model 4	30	0.885	0.243	0.877

free energy of PCBs, M05-2X and M06-2X might not be as well suited as B3LYP and HF. In addition, the results of M06-2X were superior to those of M05-2X, because M06-2X gave a better description of the medium-range exchange-correlation energies, which were manifested as attractive components of the non-covalent interaction (Zhao and Truhlar, 2008b).

3.3. Comparison of basis-sets

Effects of basis-sets on the prediction of ΔG_{OA} were analyzed while keeping the calculation method the same (Fig. 4). Based on the value of R^2, the order of basis-sets was MIDI!6D > cc-pVTZ > 6-31G(d,p) ≈ 6-31G(d) > 6-31+G(d,p). The MIDI!6D basis-set was developed to provide a well-balanced and economical basis-set that gives reasonably good molecular geometries at the Hartree-Fork level (Easton et al., 1996). Furthermore, it has been found that the MIDI!6D basis-set performed much better than the more expensive 6-31G(d) basis-set in predicting molecular geometry (Easton et al., 1996). In the present study, the MIDI!6D basis-set performed better than the other four basis-sets, which might be related to the accurate prediction of the molecular geometry of PCBs.

cc-pVTZ belongs to a three-ζ basis-set, while 6-31G(d,p), 6-31G(d) and 6-31+G(d,p) are two-ζ basis-sets. Three-ζ basis-sets describe the valence electron orbits by using three Gaussian functions, while two-ζ basis-sets use two Gaussian functions. According to the theory of quantum chemistry, the larger the basis-set, the higher the accuracy of quantitative calculation (Levine, 1983). Therefore, the prediction results of cc-pVTZ could be better than those of 6-31G(d), 6-31G(d,p) and 6-31+G(d,p).

The R^2 values obtained by using 6-31G(d) and 6-31G(d,p) were almost the same. This suggested that inclusion of an additional polarization function had little effect on the energy calculation. 6-31+G(d,p) was formed by adding a dispersion function on 6-31G(d,p). It is well known that the dispersion function is more suitable to be applied to molecules with lone pair electrons, anions and other systems with obvious negative charges (Clark et al., 1983). For the neutral PCB molecules studied in this study, the results of 6-31+G(d,p) were not as good as the results obtained by using 6-31G(d,p), which indicated that the additional diffuse function was not helpful here.

3.4. Effect of dimer formation on prediction of log K_{OA}

Based on ΔG_{OA} calculated by the optimal calculation level of HF/MIDI!6D, log K_{OA} were estimated with Eq. (1). Fig. S7 shows the experimental and the predicted values of log K_{OA} of 30 PCBs. It can be seen that the predicted log K_{OA} values were lower than the corresponding experimental values, and the deviation between them increased with increasing molecular weights. Fu et al. (2016) also found that continuum solvation models underestimated log
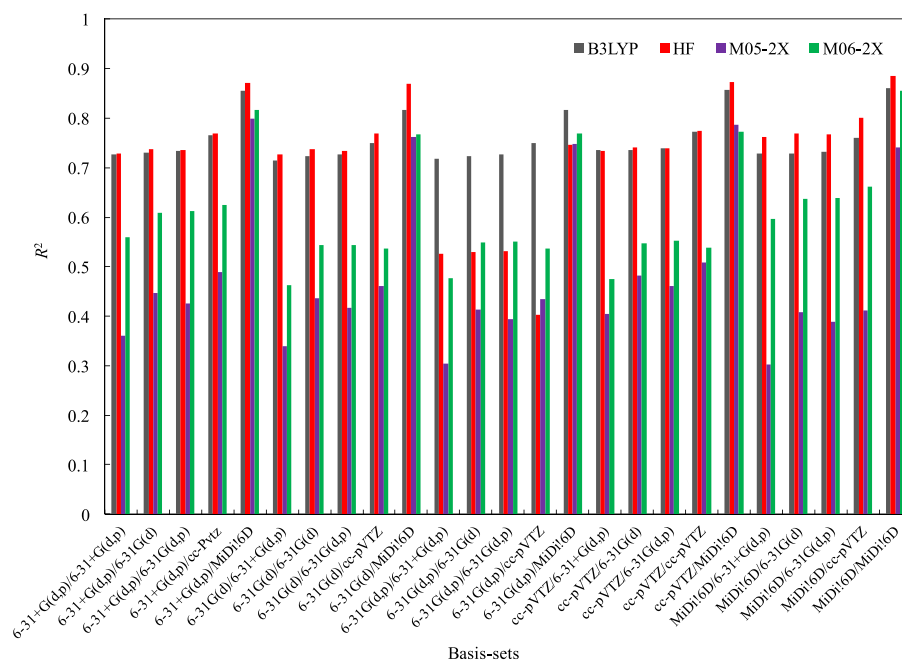
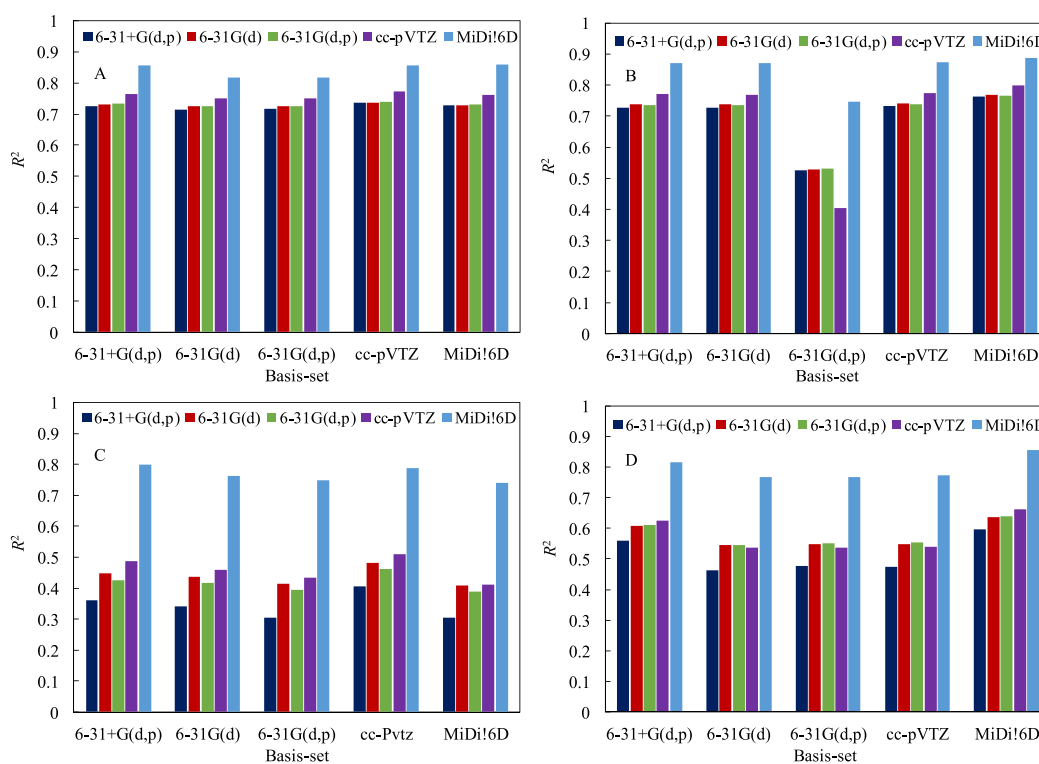**Fig. 3.** R^2 values of different theoretical methods.

Fig. 4. R^2 values of different basis-sets with the theoretical method of B3LYP(A), HF(B), M05-2X(C), M06-2X(D).

K_{OA} values for many compounds, and the underestimation generally increased with increasing log K_{OA} values, especially for compounds with log $K_{OA} > 5$.

Under the typical conditions of K_{OA} measurements, concentrations of PCBs are relatively high, and thus PCB molecules have a high probability to encounter each other to form dimers and polymers. Therefore, the experimental log K_{OA} values of PCBs might be the values of a mixture of monomers, dimers and polymers, rather than the actual value of individual molecules. Dimers and polymers are more hydrophobic than the monomeric molecule, which would lead to higher apparent log K_{OA} values. Therefore, the experimental log K_{OA} values could be higher than the predicted

values based on individual molecules. In the natural environment, PCBs might exist mainly in the form of monomers and dimers due to the very low PCB concentrations in the environment (Wild et al., 2008). Taking monomers and dimers into account without polymers should be desirable, and might improve the prediction of log K_{OA}. Therefore, for the prediction of log K_{OA} values of PCBs, the effects of the dimer were taken into account, as in the form of Eq. (5):

$$\log K_{OA} = monomer\%* \log K_{OA} \text{ (monomer)}$$
$$+ dimer\%* \log K_{OA} \text{ (dimer)} \qquad (5)$$

where monomer% and dimer% represent the ratios of monomeric and dimeric PCBs, and the sum of monomer% and dimer% was assumed to be 1. The geometry optimization and energy calculation of dimers of 30 PCBs were both conducted at the HF/MIDI!6D level. After the geometry optimization, edge-to-face stacking dimers were obtained for PCBs studied. Log K_{OA} values of dimeric PCBs were estimated with Eq. (1) based on ΔG_{OA} calculated. Subsequently, the experimental ratios of monomer and dimer of these PCBs were calculated according to Eq. (5). As listed in Table S5, the ratio of dimers increased with increasing molecular weight. The higher the molecular weight, the stronger the hydrophobicity of PCBs, and the higher the probability of the PCB molecules existing in the dimeric form. As more molecules were present in the dimeric form for PCBs with higher log K_{OA} values, the deviation between the experimental and the predicted values increased.

Subsequently, the contribution of electrostatic and non-electrostatic energies to the solvation free energy of monomeric and dimeric PCBs were calculated. As presented in Table S6, PCB congeners possessed similar electrostatic energies, while their absolute values of the non-electrostatic energy ($|\Delta G_S|$) gradually increased with increasing molecular weights. Therefore, the differences between the solvation free energies of PCB congeners were mainly attributed to the non-electrostatic energy.

In order to predict log K_{OA} values of PCBs for which no experimental data were available, QSAR models for the dimeric ratios of PCBs were developed based on the dimeric ratios and 11 molecular structural descriptors of 30 PCBs. Following the variable selection method presented by Ding et al. (2006), 10 kinds of QSAR models were obtained with different numbers of predictor variables (Table S7). It can be seen that QSAR 6 had the highest $R^2_{Y(cum)}$ and Q^2_{CV} values. Although 6 molecular structural descriptors were included, only 4 PLS components were extracted for the model. Therefore, QSAR 6 was chosen to predict dimer ratios of PCB congeners by means of Eq. (6). The prediction results were presented in Fig. S8, which showed that the predicted dimer ratios were in agreement with the experimental values. The leave-one-out cross-validated Q^2_{CV} was 0.779. These indicated that the model had good predictive ability and robustness, and could be used to estimate the dimer ratios of PCB congeners for which no data are available.

$$Dimer\% = 1.11 - 8.88 \times 10^{-7} CCR - 5.67 \times 10^{-3} \mu - 1.20$$
$$\times 10^{-2} CSEV - 2.25 \times 10^{-5} TE + 1.49 \times 10^{-2} \alpha$$
$$- 5.90 \times 10^{-3} CAA \qquad (6)$$

$n = 30$, $p = 6$, $A = 4$, $R^2_{X(cum)} = 0.997$, $R^2_{Y(cum)} = 0.818$, $Q^2_{CV} = 0.779$

Where p is the number of descriptor variables included in the model, A is the number of PLS components, and $R^2_{X(cum)}$ and $R^2_{Y(cum)}$ stand for the cumulative variance of all the predictor variables and dependent variable, respectively, explained by the extracted components.

With ratios of monomers and dimers of PCBs predicted by Eq. (6), the predicted log K_{OA} values were corrected with Eq. (5). As shown in Fig. 5, the correlation between the experimental and predicted log K_{OA} values significantly increased from 0.885 to 0.974 after amendment with the dimer. Therefore, the predictive ability of log K_{OA} of PCBs could be improved after taking the dimer into account.

3.5. Prediction of log K_{OA} values at different ambient temperatures

Temperature-dependent log K_{OA} values are very important for assessing the environmental behavior of persistent organic pollutants including PCBs. Log K_{OA} values at ambient temperatures ranging from 283.15 to 303.15 K are meaningful. However, ΔG_{OA} only at 298.15 K could be calculated by the SMD model. If log K_{OA} values of PCBs at other temperatures in the range of 283.15−303.15 K were estimated based on the thermodynamic relationship of Eq. (1) and the ΔG_{OA} values calculated at 298.15 K, ΔG_{OA} should be (weakly) temperature-independent in the temperature range.

After referring to various publications (Harner and Bidleman, 1996; Kömp and McLachlan, 1997; Wania et al., 2002; Zhang et al., 1999), it was found that a total of 15 PCBs had experimental log K_{OA} values at 283.15 K, 303.15 K and temperatures in the range of 283.15−303.15 K. The data were summarized in Table S9, and were used to check the temperature independence of ΔG_{OA}. Among the experimental log K_{OA} values, five data points at 298.15 K were found to be much lower than other values reported at 298.15 K, and even lower than the corresponding values at 303.15 K. These data were abnormal and therefore were not included for further analyses. Based on the experimental log K_{OA} values, ΔG_{OA} values were calculated. It was found that the relative range value (Range/Mean) of ΔG_{OA} of different PCBs ranged from 2.18% to 4.69%, with an average value of 3.61% (Table S9). These suggested that ΔG_{OA} was weakly temperature-dependent in the range of 283.15−303.15 K. Therefore, log K_{OA} values of PCBs in the temperature range of 283.15−303.15 K could be estimated with Eq. (1) and Eq. (5), based on ΔG_{OA} values calculated by the SMD model at 298.15 K.

Values of ΔG_{OA} of monomeric and dimeric PCBs were therefore calculated with the optimal calculation level of HF/MIDI!6D by the SMD model for the other 179 PCBs. By means of Eq. (1), log K_{OA}
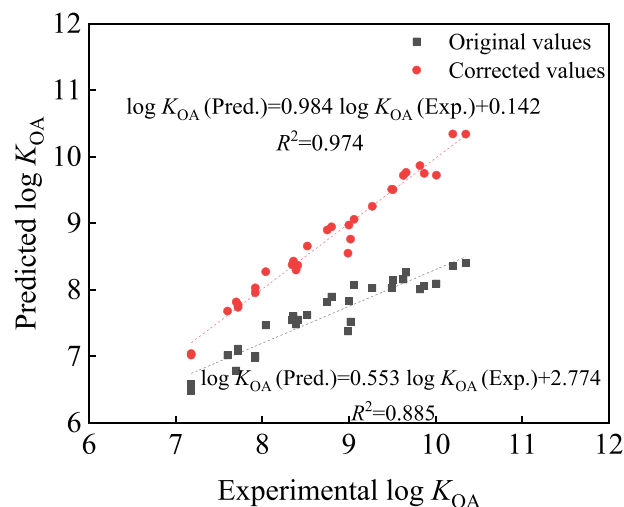

log K_{OA} (Pred.)=0.984 log K_{OA} (Exp.)+0.142
R^2=0.974

log K_{OA} (Pred.)=0.553 log K_{OA} (Exp.)+2.774
R^2=0.885

Fig. 5. Original and corrected plots of log K_{OA}.

Table 2
Comparison of the results of this study with others for the prediction of log K_{OA} of PCBs.

Method	n	R^2	RMSE	T (K)	Reference
K_{OW}-HLC	33	0.747	0.665	298.15	Meylan and Howard (2005)
QSPR	19	0.944	0.293	293.15	Yuan et al. (2016)
3D-QSAR	19	0.987	0.141	293.15	Chen et al. (2016)
QRSETP	87	0.979	0.185	263.15—303.15	Chen et al. (2003)
Fragment constant	53	0.958	0.205	283.15—303.15	Li et al. (2006)
Solvation free energy	30	0.974	0.194	283.15—303.15	This study

values of monomeric and dimeric PCBs at different ambient temperatures were estimated. Subsequently, monomer ratios and dimer ratios of PCBs were estimated by Eq. (6) based on the molecular structure descriptors of CCR, μ, CSEV, TE, α and CAA (Table S8). With log K_{OA} values and ratios of monomeric and dimeric PCBs, apparent log K_{OA} values of 209 PCBs at different ambient temperatures were predicted and listed in Table S10.

3.6. Comparison with results from other studies

Meylan and Howard (2005) estimated K_{OA} values of 33 PCBs from the octanol-water partition coefficient (K_{OW}) and Henry's law constant (HLC). Yuan et al. (2016) predicted K_{OA} values by a QSPR model based on generator-column derived K_{OA} values of 19 PCBs. Chen et al. (2016) established a 3D-QSAR model to predict K_{OA} values of PCBs based on 19 experimental K_{OA} values. Chen et al. (2003) developed QRSETP models which incorporated both environmental temperatures and molecular structures to predict log K_{OA} values in the temperature range of 263.15—303.15 K. In addition, Li et al. (2006) predicted K_{OA} values based on the fragment constant model at temperatures ranging from 283.15 to 303.15 K. The statistical parameters of these studies were listed in Table 2.

It can be seen from Table 2 that the R^2 of this study was comparable with those of others. The goodness-of-fit and predictive ability of the solvation free energy model developed in the present study were slightly lower than those of the 3D-QSAR model. For log K_{OA} of PCBs, the predictions generated by this study agreed with those of Chen et al. (2003), Li et al. (2006) and Meylan and Howard (2005), while were lower than those of Chen et al. (2016) and Yuan et al. (2016) for PCBs with high molecular weights (Fig. S9). For the solvation free energy method, the prediction of log K_{OA} values is based on the general thermodynamic relationship of Eq. (1) and ΔG_{OA} calculated with quantum chemistry. If ΔG_{OA} values of PCBs could be accurately calculated, log K_{OA} could theoretically be accurately estimated based on the general thermodynamic relationship. Therefore, the solvation free energy method has a universal application domain and is relatively independent of experimental data. However, 3D-QSAR model was more dependent on the experimental log K_{OA} values and had a limited application domain. The QSPR, QRSETP and fragment constant models showed comparable goodness-of-fit and predictive ability as this study, but they also had limited application domains. In addition, the fragment constant model was insensitive to isomeric compounds, and the predictive process of fragment constant model was complex due to the segmentation of fragments. For the K_{OW}-HLC method, it had the lowest goodness-of-fit and predictive ability. Overall, given similar goodness-of-fit and predictive ability, the solvation free energy model was superior to other models due to the universal application domain and its relative independency of experimental data. It is therefore to be concluded that the solvation free energy method presented in this study is a promising method for predicting log K_{OA} of PCBs at different ambient temperatures.

4. Conclusions

In the present study, a predictive model for log K_{OA} values of PCBs was developed based on ΔG_{OA} and the dimer ratio. For the calculation of ΔG_{OA}, HF/MIDI!6D was identified to be the optimal calculation level for both geometry optimization and energy calculation after several abnormal points were removed. Dimers could affect the partition behavior of PCBs between air and environmental organic phases, and the goodness-of-fit, predictive ability, and robustness of the prediction model were significantly improved after taking the dimer into account. Based on the predicted log K_{OA} values and ratios of monomeric and dimeric PCBs, apparent log K_{OA} values of 209 PCBs at different ambient temperatures ranging from 283.15 to 303.15 K were predicted. Compared with other reported models, the solvation free energy model had not only comparable goodness-of-fit and predictive ability, but also a universal application domain and the relative independency of experimental data. Therefore, the solvation free energy method could be a promising method for the prediction of log K_{OA}.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chemosphere.2019.125246.

References

Arunan, E., Gutowsky, H.S., 1993. The rotational spectrum, structure and dynamics of a benzene dimer. J. Chem. Phys. 98, 4294—4296.

Abraham, M.H., Al-Hussaini, A.J.M., 2005. Solvation parameters for the 209 PCBs: calculation of physicochemical properties. J. Environ. Monit. 7, 295—301.

Chakraborty, T., Lim, E.C., 1993. Study of van der Waals clusters of anthracene by laser-induced fluorescence in a supersonic jet: evidence for two structurally different dimers. J. Phys. Chem. 97, 11151—11153.

Chen, J.W., Harner, T., Schramm, K.W., Quan, X., Xue, X.Y., Kettrup, A., 2003. Quantitative relationships between molecular structures, environmental temperatures and octanol-air partition coefficients of polychlorinated biphenyls. Comput. Biol. Chem. 27, 405—421.

Chen, J.W., Harner, T., Ding, G.H., Quan, X., Schramm, K.W., Kettrup, A., 2004. Universal predictive models on octanol-air partition coefficients at different temperatures for persistent organic pollutants. Environ. Toxicol. Chem. 23, 2309—2317.

Chen, J.W., Xue, X.Y., Schramm, K.W., Quan, X., Yang, F.L., Kettrup, A., 2002. Quantitative structure-property relationships for octanol-air partition coefficients of polychlorinated biphenyls. Chemosphere 48, 535—544.

Chen, Y., Cai, X.Y., Jiang, L., Li, Y., 2016. Prediction of octanol-air partition coefficients for polychlorinated biphenyls (PCBs) using 3D-QSAR models. Ecotoxicol. Environ. Saf. 124, 202—212.

Clark, T., Chandrasekhar, J., Spitznagel, G.W., Schleyer, P.V.R., 1983. Efficient diffuse function-augmented basis sets for anion calculations. III. The 3-21+G basis set for first-row elements. Li-F. J. Comput. Chem. 4, 294—301.

Cramer, C.J., Truhlar, D.G., 2008. A universal approach to solvation modeling. Acc. Chem. Res. 41, 760—768.

Ding, G.H., Chen, J.W., Qiao, X.L., Huang, L.P., Lin, J., Chen, X.Y., 2006. Quantitative

relationships between molecular structures, environmental temperatures and solid vapor pressures of PCDD/Fs. Chemosphere 62, 1057−1063.

Doxtader, M.M., Mangle, E.A., Bhattacharya, A.K., Cohen, S.M., Topp, M.R., 1986. Spectroscopy of benzene complexes with perylene and other aromatic species. Chem. Phys. 101, 413−427.

Easton, R.E., Giesen, D.J., Welch, A., Cramer, C.J., Truhlar, D.G., 1996. The MIDI! basis set for quantum mechanical calculations of molecular geometries and partial charges. Theor. Chem. Acc. 93, 281−301.

Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G.A., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H.P., Izmaylov, A.F., Bloino, J., Zheng, G., Sonnenberg, J.L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery Jr., J.A., Peralta, J.E., Ogliaro, F., Bearpark, M., Heyd, J.J., Brothers, E., Kudin, K.N., Staroverov, V.N., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, J.C., Iyengar, S.S., Tomasi, J., Cossi, M., Rega, N., Millam, J.M., Klene, M., Knox, J.E., Cross, J.B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R.E., Yazyev, O., Austin, A.J., Cammi, R., Pomelli, C., Ochterski, J.W., Martin, R.L., Morokuma, K., Zakrzewski, V.G., Voth, G.A., Salvador, P., Dannenberg, J.J., Dapprich, S., Daniels, A.D., Farkas, Ö., Foresman, J.B., Ortiz, J.V., Cioslowski, J., Fox, D.J., 2009. Gaussian 09, Revision E.01. Gaussian, Inc., Wallingford CT, 2009.

Fu, Z.Q., Chen, J.W., Li, X.H., Wang, Y.N., Yu, H.Y., 2016. Comparison of prediction methods for octanol-air partition coefficients of diverse organic compounds. Chemosphere 148, 118−125.

Gonzalez, C., Lim, E.C., 2000. A quantum chemistry study of the van der Waals dimers of benzene, naphthalene, and anthracene: crossed (D_{2d}) and parallel-displaced (C_{2h}) dimers of very similar energies in the linear polyacenes. J. Phys. Chem. A 104, 2953−2957.

Guo, X.J., Jin, X., Lv, X.F., Pu, Y.Y., Bai, F., 2015. Real-time visualization of perylene nanoclusters in water and their partitioning to graphene surface and macrophage cells. Environ. Sci. Technol. 49, 7926−7933.

Gupta, M., da Silva, E.F., Svendsen, H.F., 2012. Modeling temperature dependency of amine basicity using PCM and SM8T implicit solvation model. J. Phys. Chem. B 116, 1865−1875.

Harner, T., Bidleman, T.F., 1996. Measurements of octanol-air partition coefficients for polychlorinated biphenyls. J. Chem. Eng. Data 41, 895−899.

Harner, T., Mackay, D., 1995. Measurement of octanol-air partition coefficients for chlorobenzenes, PCBs, and DDT. Environ. Sci. Technol. 29, 1599−1606.

Hohenstein, E.G., Chill, S.T., Sherrill, C.D., 2008. Assessment of the performance of the M05-2X and M06-2X exchange-correlation functionals for noncovalent interactions in biomolecules. J. Chem. Theory Comput. 4, 1996−2000.

Hsieh, C.M., Sandler, S.I., Lin, S.T., 2010. Improvements of COSMO-SAC for vapor-liquid and liquid-liquid equilibrium predictions. Fluid Phase Equilib. 297, 90−97.

Hunter, C.A., Sanders, J.K.M., 1990. The nature of π-π interactions. J. Am. Chem. Soc. 112, 5525−5534.

Hwang, J., Dial, B.E., Li, P., Kozik, M.E., Smith, M.D., Shimizu, K.D., 2015. How important are dispersion interactions to the strength of aromatic stacking interactions in solution? Chem. Sci. 6, 4358−4364.

Janda, K.C., Hemminger, J.C., Winn, J.S., Novick, S.E., Harris, S.J., Klemperer, W., 1975. Benzene dimer: a polar molecule. J. Chem. Phys. 63, 1419−1421.

Kendall, R.A., Dunning Jr., T.H., Harrison, R.J., 1992. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. J. Chem. Phys. 96, 6796−6806.

Kholod, Y.A., Gryn'ova, G., Gorb, L., Hill, F.C., Leszczynski, J., 2011. Evaluation of the dependence of aqueous solubility of nitro compounds on temperature and salinity: a COSMO-RS simulation. Chemosphere 83, 287−294.

Klamt, A., 2005. COSMO-RS: from Quantum Chemistry to Fluid Phase Thermodynamics and Drug Design. Elsevier Science, Ltd., Amsterdam, The Netherlands.

Klamt, A., 2011. The COSMO and COSMO-RS solvation models. WIREs: Comput. Mol. Sci. 1, 699−709.

Kömp, P., McLachlan, M.S., 1997. Octanol/air partitioning of polychlorinated biphenyls. Environ. Toxicol. Chem. 16, 2433−2437.

Lee, H., Dehez, F., Chipot, C., Lim, H.K., Kim, H., 2019. Enthalpy-entropy interplay in π-stacking interaction of benzene dimer in water. J. Chem. Theory Comput. 15, 1538−1545.

Levine, I.N., 1983. Quantum Chemistry. the United States of America, third ed. Allyn and Bacon, Inc., Rockleigh, New Jersey, pp. 436−441.

Li, Q.Q., Chen, B.L., 2014. Organic pollutant clustered in the plant cuticular membranes: visualizing the distribution of phenanthrene in leaf cuticle using two-photon confocal scanning laser microscopy. Environ. Sci. Technol. 48, 4774−4781.

Li, X.H., Chen, J.W., Zhang, L., Qiao, X.L., Huang, L.P., 2006. The fragment constant method for predicting octanol-air partition coefficients of persistent organic pollutants at different temperatures. J. Phys. Chem. Ref. Data 35, 1365−1384.

Lin, S.T., Sandler, S.I., 2002. A priori phase equilibrium prediction from a segment contribution solvation model. Ind. Eng. Chem. Res. 41, 899−913.

Marenich, A.V., Cramer, C.J., Truhlar, D.G., 2009. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. J. Phys. Chem. B 113, 6378−6396.

Marenich, A.V., Cramer, C.J., Truhlar, D.G., 2013. Generalized Born solvation model SM12. J. Chem. Theory Comput. 9, 609−620.

Meylan, W.M., Howard, P.H., 2005. Estimating octanol-air partition coefficients with octanol-water partition coefficients and Henry's law constants. Chemosphere 61, 640−644.

Miertus, S., Scrocco, E., Tomasi, J., 1981. Electrostatic interaction of a solute with a continuum. A direct utilization of ab initio molecular potentials for the prevision of solvent effects. Chem. Phys. 55, 117−129.

Miliordos, E., Aprà, E., Xantheas, S.S., 2014. Benchmark theoretical study of the π-π binding energy in the benzene dimer. J. Phys. Chem. A 118, 7568−7578.

Petersson, G.A., Al-Laham, M.A., 1991. A complete basis set model chemistry. II. Open-shell systems and the total energies of the first-row atoms. J. Chem. Phys. 94, 6081−6090.

Petersson, G.A., Bennett, A., Tensfeldt, T.G., Al-Laham, M.A., Shirley, W.A., Mantzaris, J., 1988. A complete basis set model chemistry. I. The total energies of closed-shell atoms and hydrides of the first-row atoms. J. Chem. Phys. 89, 2193−2218.

Phillips, K.L., Di Toro, D.M., Sandler, S.I., 2011. Prediction of soil sorption coefficients using model molecular structures for organic matter and the quantum mechanical COSMO-SAC model. Environ. Sci. Technol. 45, 1021−1027.

Piacenza, M., Grimme, S., 2005. Van der Waals complexes of polar aromatic molecules: unexpected structures for dimers of azulene. J. Am. Chem. Soc. 127, 14841−14848.

Shen, H.Q., Chen, Z.M., Li, H., Qian, X., Qin, X., Shi, W.X., 2018. Gas-particle partitioning of carbonyl compounds in the ambient atmosphere. Environ. Sci. Technol. 52, 10997−11006.

Sherrill, C.D., 2013. Energy component analysis of π interactions. Acc. Chem. Res. 46, 1020−1028.

Stephens, P.J., Devlin, F.J., Chabalowski, C.F., Frisch, M.J., 1994. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. J. Phys. Chem. 98, 11623−11627.

Stewart, J.J.P., 2013. Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. J. Mol. Model. 19, 1−32.

Thompson, J.D., Cramer, C.J., Truhlar, D.G., 2004. New universal solvation model and comparison of the accuracy of the SM5.42R, SM5.43R, C-PCM, D-PCM, and IEF-PCM continuum solvation models for aqueous and organic solvation free energies and for vapor pressures. J. Phys. Chem. A 108, 6532−6542.

Tirado-Rives, J., Jorgensen, W.L., 2008. Performance of B3LYP density functional methods for a large set of organic molecules. J. Chem. Theory Comput. 4, 297−306.

Wania, F., Lei, Y.D., Harner, T., 2002. Estimating octanol-air partition coefficients of nonpolar semivolatile organic compounds from gas chromatographic retention times. Anal. Chem. 74, 3476−3483.

White, R.P., Niesse, J.A., Mayne, H.R., 1998. A study of genetic algorithm approaches to global geometry optimization of aromatic hydrocarbon microclusters. J. Chem. Phys. 108, 2208−2218.

Wild, E., Cabrerizo, A., Dachs, J., Jones, K.C., 2008. Clustering of nonpolar organic compounds in lipid media: evidence and implications. J. Phys. Chem. A 112, 11699−11703.

Yuan, J.T., Yu, S.L., Zhang, T., Yuan, X.J., Cao, Y.Y., Yu, X.C., Yang, X., Yao, W., 2016. QSPR models for predicting generator-column-derived octanol/water and octanol/air partition coefficients of polychlorinated biphenyls. Ecotoxicol. Environ. Saf. 128, 171−180.

Zhang, X.M., Schramm, K.-W., Henkelmann, B., Klimm, C., Kaune, A., Kettrup, A., Lu, P., 1999. A method to estimate the octanol-air partition coefficient of semivolatile organic compounds. Anal. Chem. 71, 3834−3838.

Zhao, Y., Truhlar, D.G., 2008a. Density functionals with broad applicability in chemistry. Acc. Chem. Res. 41, 157−167.

Zhao, Y., Truhlar, D.G., 2008b. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. Theor. Chem. Acc. 120, 215−241.

Zhao, Y., Schultz, N.E., Truhlar, D.G., 2006. Design of density functionals by combining the method of constraint satisfaction with parametrization for thermochemistry, thermochemical kinetics, and noncovalent interactions. J. Chem. Theory Comput. 2, 364−382.