Modelling the octanol-air partition coefficient of aromatic pollutants based on the solvation free energy and the dimer effect

Li, W.; Chen, D.; Chen, S.; Zhang, J.; Song, G.; Shi, Y.; ... ; Peijnenburg, W.J.G.M.

Chemosphere 309 (2022) 136608

Contents lists available at ScienceDirect

Chemosphere

journal homepage: www.elsevier.com/locate/chemosphere

Modelling the octanol-air partition coefficient of aromatic pollutants based on the solvation free energy and the dimer effect

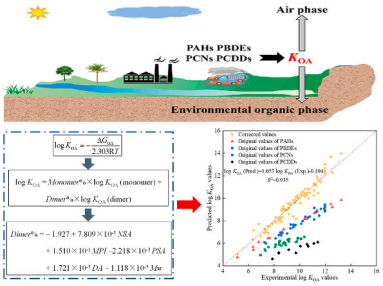
Wanran Li [a], Dezhi Chen [a], Shuhua Chen [b,**], Jing Zhang [b], Guobin Song [a], Yawei Shi [a], Ya Sun [a], Guanghui Ding [a,*], Willie J.G.M. Peijnenburg [c,d]

[a] College of Environmental Science and Engineering, Dalian Maritime University, Linghai Road 1, Dalian, 116026, PR China
[b] College of Environmental and Chemical Engineering, Dalian University, Dalian, 116622, 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

- Generalized predictive models were developed for K_{OA} of aromatic pollutants.
- The underlying π-π interactions for the formation of dimers were dissected.
- A universal QSAR for dimer percentages was developed based on the π-π interactions.
- The dimer effect on the estimation of log K_{OA} was analyzed and amended.

GRAPHICAL ABSTRACT



ARTICLE INFO

Handling Editor: Keith Maruya

Keywords:
Octanol-air partition coefficient
Aromatic pollutants
Solvation free energy
Dimer
π-π interactions

ABSTRACT

In this study, generalized predictive models were developed to estimate K_{OA} of four kinds of aromatic pollutants based on the calculated solvation free energy and taking the dimer effect into account. Uncorrected log K_{OA} values, which were directly estimated from the calculated solvation free energy of individual molecules, underestimated experimental values, and the deviation increased with increasing log K_{OA}. Dimers were found to greatly affect the apparent K_{OA} values of these aromatic pollutants, which were driven by π-π interactions. London dispersion and exchange-repulsion terms were identified to be dominant components of the underlying π-π interactions. It is interesting to find that the π-π interactions of polybrominated diphenyl ethers correlate with not only the molecular polarizability but also the size of opposing aromatic surfaces, which leads to a different trend of π-π interactions from other aromatic pollutants. A universal quantitative structure-activity relationship model was developed to estimate the proportion of dimers based on five molecular structural descriptors relevant to the π-π interactions. After calibration with the dimer effect, estimations of log K_{OA} were consistent with experimental values. Therefore, the dimer effect should be taken into consideration when investigating the partition behavior of aromatic pollutants, and the solvation free energy model could be an alternative method for the prediction of K_{OA}.

* Corresponding author.
** Corresponding author.
E-mail addresses: chenshuhua@dlu.edu.cn (S. Chen), guanghuiding@dlmu.edu.cn (G. Ding).

https://doi.org/10.1016/j.chemosphere.2022.136608
Received 28 June 2022; Received in revised form 20 September 2022; Accepted 23 September 2022
Available online 29 September 2022

1. Introduction

Polycyclic aromatic hydrocarbons (PAHs), polybrominated diphenyl ethers (PBDEs), polychlorinated naphthalenes (PCNs), and polychlorinated dibenzo-p-dioxins (PCDDs) are ubiquitous aromatic pollutants, which can be accumulated and biomagnified through the food chain and consequently lead to adverse toxicological effects (Atkinson et al., 2018; Viluksela and Pohjanvirta, 2019). At present, these pollutants are still being released as unintended by-products from combustion sources, industrial sources, and vehicle exhaust sources. Correspondingly, they have been detected in different environmental matrices throughout the world (Wang et al., 2019). The widespread presence of these pollutants causes a severe concern due to their persistence, bioaccumulation, long-range transport potential and significant toxicity. Therefore, it is of great importance to investigate their environmental fate and behavior.

Among many physicochemical properties, octanol-air partition coefficient (K_{OA}) is a key parameter determining the partition of organic pollutants between the gaseous phase and environmental organic phases, such as soil, vegetation and aerosols. K_{OA} is also a key descriptor of the mobility and long-range transport potential of organic pollutants in the global environment. Furthermore, K_{OA} is closely related to other parameters, such as the gas-particle partition coefficient, the octanol-water partition coefficient (K_{OW}) and the Henry's law constant (*HLC*), and could be used to describe these parameters (Harner, 1998; Xu et al., 2014). As the experimental determination of K_{OA} values of all of these organic pollutants is costly, laborious and unreliable, predictive models are desirable to estimate the K_{OA} values.

In previous studies, a number of predictive models have been developed for K_{OA}, such as the direct calculation method based on K_{OW} and *HLC* (Meylan and Howard, 2005), the conductor-like screening model for realistic solvents (COSMO-RS) (Parnis et al., 2015), quantitative structure-activity relationships (QSARs) (Jin et al., 2017; Wang et al., 2017), and the thermodynamic method based on the solvation free energy (Fu et al., 2016; Nedyalkova et al., 2019). It is noteworthy that the accuracy of the direct calculation method is highly dependent on the accuracy of K_{OW} and *HLC* values, while the QSAR models are often limited by their application domains. While the solvation free energy model is a relatively universal model without restrictions on the application domain. Based on the typical thermodynamic relationship of Eq. (1), K_{OA} values of organic pollutants can be directly estimated from the calculated solvation free energy transferred from the gaseous phase to the octanol phase (ΔG_{OA}, kcal/mol).

$$\log K_{OA} = -\frac{\Delta G_{OA}}{2.303RT} \quad (1)$$

where R [8.314 J/(mol·K)] is the gas constant, and T is the absolute temperature (K).

Many continuum solvation models, such as the SMx model, the COSMO-RS model and the polarizable continuum model (PCM), have been developed to calculate the solvation free energy (Miertus et al., 1981; Thompson et al., 2004; Klamt, 2005; Cramer and Truhlar, 2008; Marenich et al., 2009, 2013; Kholod et al., 2011; Gupta et al., 2012). Among them, the Solvation Model Density (SMD) model (Marenich et al., 2009) is one of the most precise continuum solvation models due to its consideration of the first solvation shell. In our previous study, it was found that ΔG_{OA} values of polychlorinated biphenyls (PCBs) calculated by the SMD model based on the HF/MiDi!6D level were in accordance with the experimental values (Li et al., 2020). However, it is not known if it can be generalized to other organic pollutants. Therefore, in the present study, the SMD model combined with the HF/MiDi!6D level was also employed to calculate ΔG_{OA} values of PAHs, PBDEs, PCNs and PCDDs, which were subsequently used to estimate their log K_{OA}.

It is well known that, driven by the π-π interactions, organic chemicals with an aromatic π-system in their molecular structures, such as benzenes and PAHs, often form dimers and polymers (Doxtader et al., 1986; Hunter and Sanders, 1990; Arunan and Gutowsky, 1993; Chakraborty and Lim, 1993; Sato et al., 2005; Podeszwa and Szalewicz, 2008; Miliordos et al., 2014). As PCBs, PBDEs, PCNs and PCDDs have aromatic π-systems in their molecular structures, theoretically they could also form the corresponding dimers and polymers. During the experimental measurement of K_{OA}, concentrations of these chemicals are relatively higher than those in the environment, and thus their molecules have higher probability to encounter each other and to form more dimers and polymers. Recently, the formation of dimers and few trimers of typical PAH, PCB, PCN, PBDE and PCDD in the *n*-octanol phase has been supported by results of the molecular dynamics simulation (Li et al., 2022). Given the higher lipophilicity, the presence of more aggregates is expected to raise the apparent K_{OA} values relative to the theoretically predictive values based on monomers only. As for dimers and polymers, it has been showed that taking dimers into account without polymers could effectively improve the prediction of K_{OA} of PCBs already (Li et al., 2020). Therefore, it is desirable to consider the effect of dimers when K_{OA} values of these aromatic pollutants are estimated, which could not only improve the prediction of K_{OA}, but also simplify the calculation and analyses.

The formation of dimers of aromatic chemicals is driven by π-π interactions, the magnitude and orientation of which depend on molecular structure, molecular weight, and the size of the opposing aromatic surfaces (Hunter and Sanders, 1990; Sherrill, 2013; Hwang et al., 2015; Abdalla and Fink, 2016). Substituents connected to the aromatic ring also affect π-π interactions, as substituents could donate or withdraw electron density and consequently change the electrostatic interaction (Hunter and Sanders, 1990; Sherrill, 2013). Therefore, the strength of π-π interactions is different for aromatic chemicals with different molecular structures, and the proportion of dimers should be different. As the symmetry-adapted perturbation theory (SAPT) could dissect the π-π interactions into four components, including electrostatics term, London dispersion term, induction (polarization) term and exchange-repulsion term (Szalewicz, 2012), it can be used to explore the major contribution to the π-π interactions. In addition, QSAR models could be developed to estimate the proportion of dimers for these aromatic pollutants based on molecular structural descriptors highly relative to the π-π interactions.

Therefore, in this study, ΔG_{OA} values of aromatic pollutants, including PAHs, PBDEs, PCNs and PCDDs, were calculated by using the SMD model, and log K_{OA} values were estimated based on the thermodynamic relationship presented in Eq. (1). The effect of dimer formation on the estimation of log K_{OA} was analyzed, and the underlying π-π interactions for the formation of dimers was explored. Subsequently, a universal QSAR model was developed to estimate the proportion of dimers according to the π-π interactions, and the dimer effect was taken into consideration to correct the predictive log K_{OA} values.

2. Materials and methods

2.1. Data set

Experimental log K_{OA} values of 16 PAHs, 30 PBDEs, 24 PCNs and 10 PCDDs at 298.15 K were collected from previous publications (Harner and Bidleman, 1998; Mackay and Callcott, 1998; Harner et al., 2000; Treves et al., 2001; Harner and Shoeib, 2002; Wania et al., 2002; Odabasi et al., 2006). For chemicals with more than one experimental log K_{OA} value, average values were adopted. All of these log K_{OA} values are listed in Table S1.

2.2. Computational methods

Molecular structures of both monomeric and dimeric aromatic pollutants studied were drawn by ChemBioDraw Ultra (Version 14.0, CambridgeSoft Corporation). For the dimers, the initial conformations

were generated as the edge-to-face stacking type or offset stacking type following the results of the molecular dynamic simulation on the aggregation behavior of these aromatic pollutants. The calculation process of ΔG_{OA} with the SMD model involved three steps: geometry optimization, frequency calculation and energy calculation, which were all performed by using Gaussian 09-E01 (Frisch et al., 2009). The HF/MiDi!6D level, which has been chosen as the optimal calculation level to calculate ΔG_{OA} values of PCBs in our previous study (Li et al., 2020), was also adopted to calculate ΔG_{OA} in this study. IBM SPSS Statistics 22.0 software (IBM Corp.) was used to develop linear regression models between the experimental and predicted ΔG_{OA} values.

For the calculation of π-π interactions, the third generation of density functional dispersion correction (DFT-D3) (Grimme et al., 2010) with Becke-Johnson (BJ) damping (Grimme et al., 2011) was employed, which could describe the dispersion effect precisely. In addition, the basis set superposition error (BSSE) was corrected by the counterpoise (CP) method proposed by Boys and Bernardi (1970). The π-π interactions of these aromatic pollutants in the octanol phase were obtained by using Eq. (2).

$$E(\text{Interaction}) = E(\text{Dimer}) - 2 \times E(\text{Monomer}) + E(\text{BSSE}) \quad (2)$$

where *E*(Interaction), *E*(Dimer), *E*(Monomer), and *E*(BSSE) represent the energy of π-π interactions, dimers, monomers and BSSE, respectively. The geometry optimization was conducted at B3LYP-D3(BJ)/6-311G(d, p) level. *E*(Dimer), *E*(Monomer) and *E*(BSSE) were calculated at the B3LYP-D3(BJ)/6-311+G(2d,p) level. With the energy component analysis of SAPT, the dominant contribution terms for the π-π interactions of these aromatic pollutants were explored.

2.3. Estimation of the proportion of monomeric and dimeric aromatic pollutants

It had been supposed that experimental log K_{OA} values of these aromatic pollutants were the sum of corresponding values of monomers and dimers. Therefore the percentages of monomers and dimers of aromatic pollutants with experimental log K_{OA} values, were estimated based on the experimental log K_{OA} values and predicted values of corresponding monomers and dimers. Subsequently, a universal QSAR model was developed based on these percentages obtained, which were used to estimate the dimer percentages of aromatic pollutants without experimental log K_{OA} values. A total of 15 molecular structural descriptors were selected to develop the QSAR models as these descriptors are closely related to the π-π interactions (Politzer et al., 2001; Byrd and Rice, 2006). They are dihedral angle (*DA*), molecular weight (*Mw*), molecular volume, density, minimal electrostatic potential (*ESP*) value, maximum *ESP* value, overall surface area, positive *ESP* surface area, negative *ESP* surface area, overall average *ESP* value, positive average *ESP* value, negative average *ESP* value, molecular polarity index (*MPI*) (Liu et al., 2021), nonpolar surface area (*NSA*) and polar surface area (*PSA*). *DA* was calculated by using GaussView 6.0, *Mw* was calculated by the PM7 method in MOPAC 2016 (Stewart Computational Chemistry, Colorado Springs, CO, USA), and other molecular structural descriptors were obtained from the quantitative analyses of the molecular surface module (Lu and Chen, 2012a) implemented in Multiwfn software (Lu and Chen, 2012b). QSAR models were developed by using the partial least-squares (PLS) regression in Simca-S (Version 13.0, Umetri AB & Erisoft AB) software following the method of Ding et al. (2006).

2.4. Statistical analyses

The goodness-of-fit, predictive accuracy and robustness of linear regression models between the experimental and predicted ΔG_{OA} were evaluated by the coefficient of determination (R^2), the root mean square error (*RMSE*) and the leave-one-out cross validation statistic (Q^2_{CV}). The *F*-test and *t*-test were performed to test whether the linear regression model and the regression coefficients were significant at a significant level of 0.05.

The performance of a PLS model was characterized by the number of PLS principal components (*A*), the cumulative variance of all the predictor variables and the response variables explained by the extracted components ($R^2_{X(cum)}$ and $R^2_{Y(cum)}$), the cumulative cross-validation coefficient ($Q^2_{(cum)}$), and *RMSE*. Variable Importance in the Projection (*VIP*) is a parameter that shows the importance of a variable in a model. A predictor variable with a *VIP* value higher than 1 is considered to be more relevant for explaining the response variable.

In addition, the studentized residual (r_i) and leverage value (p_{ii}) were used to identify abnormal data points as outliers and high leverage points, respectively. The values of r_i and p_{ii} can be calculated with Eq. (3) and Eq. (4), respectively.

$$r_i = \frac{e_i}{\widehat{\sigma}_{(i)}\sqrt{1-p_{ii}}} \quad (3)$$

$$p_{ii} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2} \quad (4)$$

where e_i is the residual error of i, $\widehat{\sigma}_{(i)}$ is the estimate of standard deviation σ, x_i is the observed value of i, and $\bar{x}$ is the average of the observed values. If a data point has the studentized residual that is larger than 3 (in absolute value), it is identified to be an outlier. If the leverage value of a data point is more than three times greater than the mean leverage value, it is identified as a high leverage point.

3. Results and discussion

3.1. Development and validation of ΔG_{OA} models for typical aromatic pollutants

As experimental log K_{OA} values were available for 16 PAHs, 30 PBDEs, 24 PCNs and 10 PCDDs, experimental ΔG_{OA} values of these aromatic pollutants were obtained according to Eq. (1). ΔG_{OA} of these aromatic pollutants were also calculated with the SMD model at the HF/MiDi!6D level. All experimental log K_{OA} values, as well as the experimental and calculated ΔG_{OA} values of these aromatic pollutants were listed in Table S1. Based on these values, simple linear regression models were developed, named Model 1, Model 2, Model 3 and Model 4 for PAHs, PBDEs, PCNs and PCDDs, respectively.

From Figs. S1, S2, S3 and S4, it can be seen that the calculated ΔG_{OA} values are in good agreement with corresponding experimental values with high R^2 and Q^2_{CV}, and low *RMSE* for Models 1–4. The equivariance and normal distribution of residuals were verified by the standardized residual scatter plots, standard residual histograms and normal probability plots. With regard to the *F*-test, $F(1844.868) > F_\alpha(1, 14)$, $F(592.162) > F_\alpha(1, 28)$, $F(94.601) > F_\alpha(1, 22)$ and $F(56.809) > F_\alpha(1, 8)$ were obtained for Models 1–4, respectively. As for the *t*-test, $|t| \geq t_{\alpha/2}$ was obtained for all of these models. It can also be seen that none of the data points were identified as outliers or high leverage points. Therefore, these regression models were well-fitted and statistically significant based on the results presented above. However, systematic overestimations were noticed for the calculated ΔG_{OA} values, which could lead to systematic underestimations of log K_{OA} values.

3.2. Effect of the dimer on the estimation of log K_{OA} and the underlying π-π interactions

Based on the ΔG_{OA} values calculated, log K_{OA} values of typical PAHs, PBDEs, PCNs and PCDDs were estimated with Eq. (1). The experimental and estimated log K_{OA} values of these aromatic pollutants are shown in Fig. 1A. It can be seen that the estimated log K_{OA} values were lower than the corresponding experimental values, and the deviations increased with increasing experimental values. In our previous study, log K_{OA}

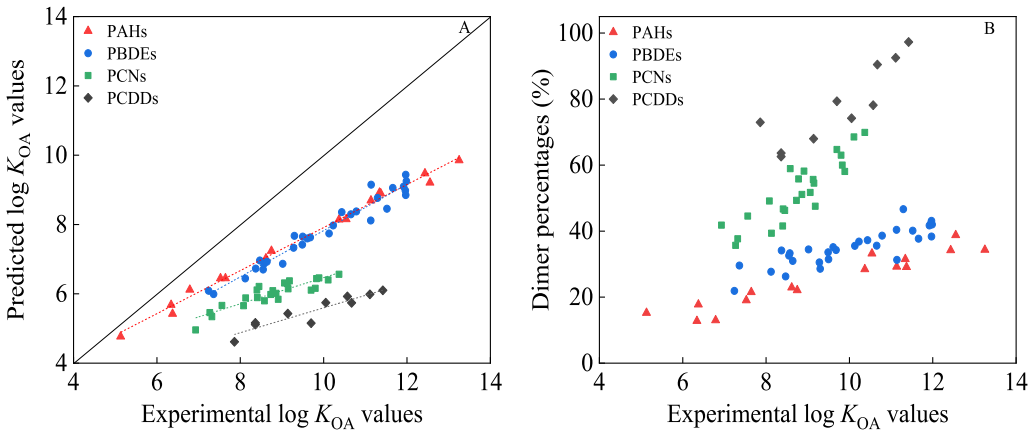


Fig. 1. Experimental and estimated log K_{OA} values (A), and dimer percentages of selected PAHs, PBDEs, PCNs and PCDDs (B).

values of PCBs estimated based on ΔG_{OA} values calculated with the SMD model, were also found to be underestimated (Li et al., 2020). Furthermore, Fu et al. (2016) also reported the underestimation of log K_{OA} values based on calculated ΔG_{OA} values with the SM8AD model, and the deviation also increased with increasing log K_{OA} values, especially for compounds with log $K_{OA} > 5$.

Driven by the π-π interactions, aromatic chemicals, such as benzenes and PAHs, often form dimers and polymers (Hunter and Sanders, 1990; Sinnokrot and Sherrill, 2004; Wild et al., 2008; Sherrill, 2013; Miliordos et al., 2014; Hwang et al., 2015). Under typical measurement conditions of K_{OA}, concentrations of these aromatic chemicals are relatively high, so that their molecules have a higher probability to encounter each other, and then form dimers and polymers during the measurement. Thus, experimental log K_{OA} values should be combined values of a mixture of monomers, dimers and polymers, instead of the values of monomers. As the lipophilicity of dimers and polymers is higher than that of monomers, the experimental log K_{OA} should be higher than the theoretically predictive values based on monomers only. This could be the reason that the estimated log K_{OA} values were lower than the corresponding experimental values as shown in Fig. 1A.

It has been proven that taking monomers and dimers into account, whilst ignoring polymers, could effectively improve the prediction of K_{OA} of PCBs already (Li et al., 2020). Therefore, the effect of dimers was taken into account in this study, and a correction equation (Eq. (5)) was adopted for the estimation of log K_{OA}.

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

where *monomer*% and *dimer*% are percentages of monomeric and dimeric aromatic chemicals, and the sum of their values for a chemical is supposed to be 100%. According to Eq. (5), the percentages were

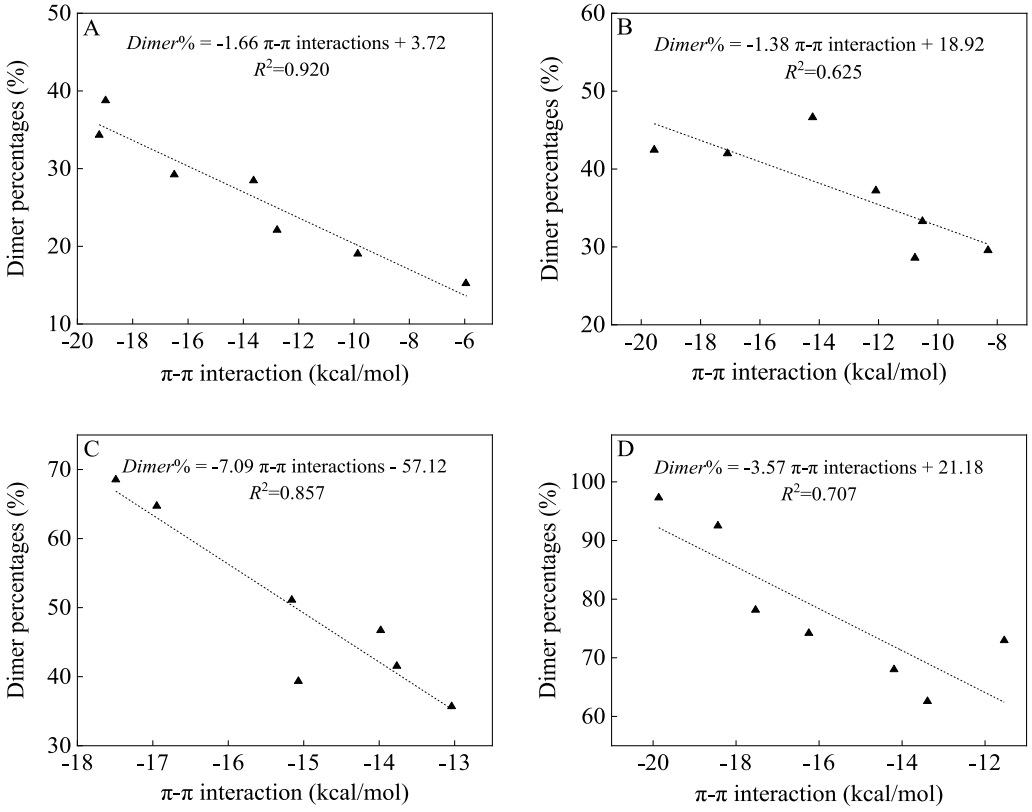


Fig. 2. The correlation between the dimer percentages and π-π interactions of selected PAHs (A), PBDEs (B), PCNs (C) and PCDDs (D) in the *n*-octanol phase.

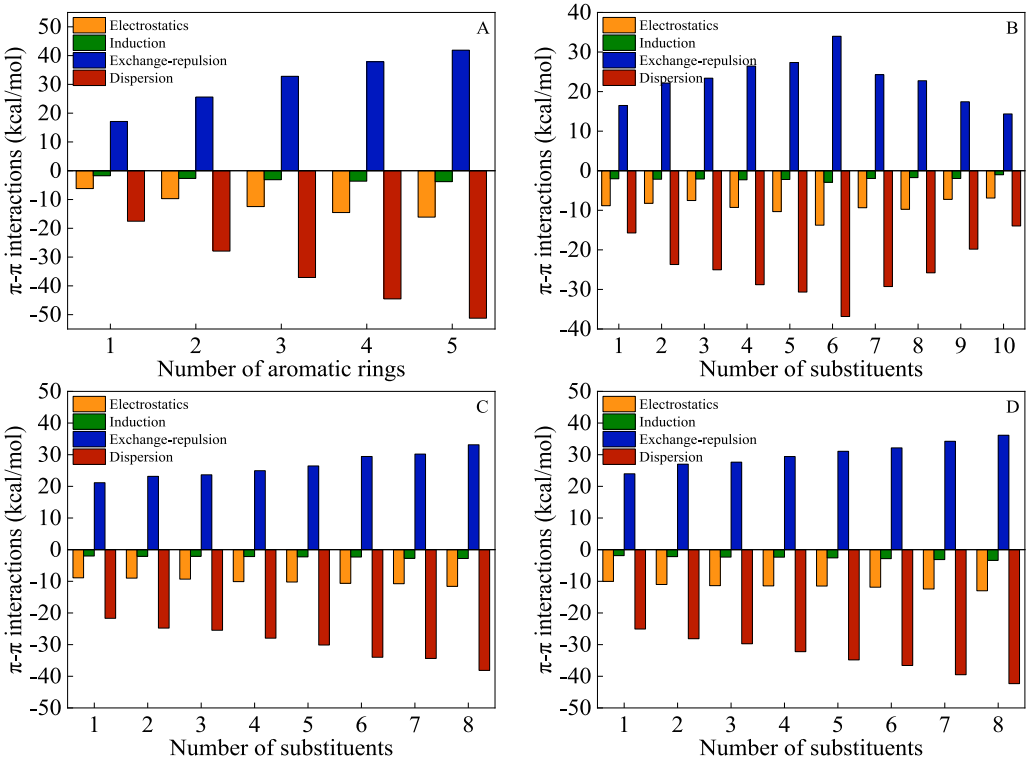


Fig. 3. The energy component analysis of the π-π interactions of selected PAHs (A), PBDEs (B), PCNs (C) and PCDDs (D).

calculated based on experimental log K_{OA} values and estimated values of corresponding monomers and dimers, the values of which were listed in Table S2. For these aromatic pollutants, the plot of dimer percentages versus experimental log K_{OA} values was shown in Fig. 1B. It can be seen that the dimer percentages are different for different aromatic pollutants with the same log K_{OA} value in the order of PCDDs > PCNs > PBDEs > PAHs. As the lipophilicity of dimers is higher than that of monomers, aromatic pollutants with more dimers have greater deviations between the experimental and estimated log K_{OA} values. From Fig. 1A, it can be seen that the order of the deviation between the experimental and estimated log K_{OA} is also PCDDs > PCNs > PBDEs > PAHs, which agrees with the above discussion.

As π-π interactions are the main driving force for the formation of dimers, dimer percentages of different aromatic pollutants should be related to the strength of π-π interactions. Therefore, π-π interactions in the octanol phase were calculated for selected aromatic pollutants at the B3LYP-D3(BJ)/6-311+G(2d,p) level with the CP mission, results of which were listed in Table S3. The strength of π-π interactions of these aromatic pollutants ranged from −5.96 to −21.50 kcal/mol. It was interesting to find that the π-π interactions calculated in the *n*-octanol phases were well correlated with the dimer percentages (Fig. 2).

Subsequently, a zeroth-order SAPT method with the jun-cc-pVDZ basis-set was adopted to analyze the major contribution to the π-π interactions. The SAPT method dissects the π-π interactions into an electrostatics term, a London dispersion term, an induction (polarization) term and an exchange-repulsion term. Among them, the electrostatics and induction terms relate to the electrostatic potential (Murray and Politzer, 2011), while the London dispersion and the exchange-repulsion terms relate to the van der Waals (vdW) potential (Lu and Chen, 2020). As π-π interactions are usually reported as negative values, the more negative the reported values, the stronger the π-π interactions. From Fig. 3, it can be seen that the London dispersion and exchange-repulsion terms were the dominant components of the π-π interactions. The electrostatics term played a relatively unimportant role, while the contribution of the induction term was almost negligible. Based on the values of these terms, the London dispersion term was found to be the main attractive term, which has also been identified as the major source of π-π interactions by Tsuzuki et al. (2002) and Sinnokrot and Sherrill (2004).

From Fig. 3, it can also be seen that the absolute values of all these energy components of PAHs, PCNs and PCDDs increased with increasing numbers of the substituent. As molecular weight increased with increasing numbers of the substituent, the changes of energy components of PAHs, PCNs and PCDDs were in accordance with the changes of their molecular weights. However, for PBDEs, the absolute values of all the energy components increased with increasing numbers of bromine substituents from 1 to 6, and then decreased from 7 to 10. It has been reported that the π-π interactions of benzene correlate highly with not only the molecular polarizability but also the size of the opposing aromatic surfaces (Zeinalipour-Yazdi and Pullman, 2006; Zhang, 2011). It is well known that the polarizability correlates well with molecular weight. Therefore, the change of the energy components might be related to the impacts of molecular weight and the size of the opposing aromatic surfaces of PBDE congeners.

The size of the opposing aromatic surfaces can be characterized by the integral domains enclosed by the reduced density gradient (RDG) isosurface (Johnson et al., 2010; Lu and Chen, 2012b; Manzetti and Lu, 2012; Lu and Manzetti, 2014; Jiao et al., 2017). Therefore, RDG isosurfaces of typical aromatic pollutants were drawn and shown in Fig. S5 by using Multiwfn (Lu and Chen, 2012b) and VMD (Humphrey et al., 1996). The green region of the RDG isosurface between two aromatic molecules was identified as the π-π interactions. The bigger the green region, the stronger the π-π interactions. Subsequently, the integral volume enclosed by the RDG isosurface was calculated by Multiwfn and listed in Table S3. Since PAHs, PCNs and PCDDs were all co-planar compounds, the volumes enclosed by their RDG isosurfaces systematically increased with increasing molecular weights. However, PBDEs are non-planar compounds, and the dihedral angle of PBDE congeners changed with the number of substituents, resulting in corresponding changes of the volume enclosed by the RDG isosurface. It was found that the volume increased with increasing numbers of the substituent for PBDEs with 1–6 substituents, while the value decreased for PBDEs with 7–10 substituents. Therefore, the π-π interactions increased firstly for

Table 1
VIP values of predictor variables and PLS weights of the QSAR model for the dimer percentage [a].

Variable	*VIP*	*w*[1]	*w*[2]	*w*[3]	*w*[4]
NSA	1.199	**0.584**	0.196	**0.651**	0.378
MPI	1.051	**−0.537**	0.328	0.345	0.410
PSA	1.011	**−0.506**	0.269	0.399	−0.406
DA	0.911	0.301	**0.823**	−0.229	−0.366
Mw	0.780	0.154	−0.323	0.494	**−0.625**

[a] The bold-faced values indicate that the PLS components are mainly loaded on the corresponding variables.

PBDEs with 1–6 bromine substituents, and then decreased for PBDEs with 7–10 bromine substituents. Furthermore, it was noticed that the dimer percentages of PBDEs with 8–10 bromine substituents cannot be obtained because of the scarcity of their experimental log K_{OA} values. Therefore, the measurement of log K_{OA} of these PBDEs is required in order to provide more basic physicochemical data for PBDEs and to verify the relationship between the π-π interactions and the number of bromine substituents.

3.3. The estimation of dimer percentages and the correction of log K_{OA}

From Fig. 2, it could be seen that the linear relationships between π-π interactions and dimer percentages for different kinds of aromatic pollutants were different. Therefore, a universal QSAR model was developed to predict the dimer percentages for these aromatic pollutants. It has been reported that the π-π interactions mainly involve two parts: the *ESP* and the vdW potential (Lu and Chen, 2020). Therefore, a total of 15 molecular structural descriptors related to *ESP* and vdW potential were selected for the modelling. 75% of experimental dimer percentages were randomly selected as the training set, and the remaining 25% experimental dimer percentages were used as the external validation set (Table S4). Following the variable selection method reported by Ding et al. (2006), five most relevant molecular structural descriptors, including *DA*, *NSA*, *PSA*, *Mw* and *MPI*, were selected at last, the values of which were listed in Table S4. The predictive model for the dimer percentage is presented as Eq. (6). Here, *DA* is a parameter identifying the planar or non-planar structure of a molecule, *NSA* and *PSA* are measurements of the vdW potential and the *ESP* respectively, and *Mw* and *MPI* relate to the polarity of a chemical.

$$Dimer\,percentages = -1.927 + 7.809\times10^{-3}NSA + 1.510\times10^{-1}MPI - 2.218\times10^{-3}PSA + 1.721\times10^{(-3)}DA - 1.118\times10^{(-3)}Mw \quad (6)$$

$N_{TRA} = 60$, $p = 5$, $A = 4$, $R^2_{X(cum)} = 0.994$, $R^2_{Y(cum)} = 0.815$, $Q^2_{(cum)} = 0.793$, $RMSE = 0.071$, $N_{VAL} = 20$, $RMSE_{VAL} = 0.062$, $Q^2_{VAL} = 0.800$. where N_{TRA} and N_{VAL} are the numbers of data points in the training and validation set respectively, and p is the number of predictor variables used for the modelling. Four PLS components were extracted, and they explained 99.4% and 81.5% of the variance of the predictor and response variables, respectively. High Q^2 and low *RMSE* suggest that the model is robust and has a good predictive ability. It can furthermore be seen from Fig. S6 that the predicted dimer percentages by the QSAR model are in agreement with the experimental percentages. This also revealed that the model has a good predictive ability.

VIP values of predictor variables and PLS weights of the QSAR model are listed in Table 1. Predictor variables with larger *VIP* values are more relevant for modelling the dimer percentage. Predictor variable that is important to the *a*th PLS component, have a larger absolute value of *w*[*a*] (Wold et al., 2001). As the *VIP* values of *NSA*, *MPI* and *PSA* are higher than 1, they are identified to be more important variables for modelling the dimer percentage. From *w*[1] values, it can be seen that the first PLS component primarily condenses information of *MPI*, *NSA* and *PSA*. These variables characterize the polarizability, vdW potential and *ESP* of these molecules, respectively. As these variables all have high *VIP* values, they are the most relevant predictors for modelling the dimer percentage. The second PLS component mainly describes information of the dihedral angle, which characterizes the molecular geometry. The third PLS component mainly condenses partial information of *NSA*. The fourth PLS component is dominated by *Mw*, and provides further information on the structural properties of these aromatic pollutants that affect the dimer percentages.

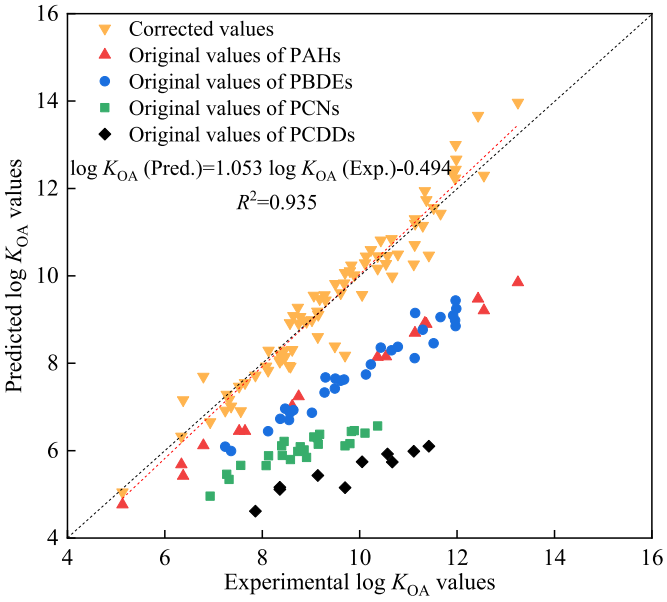


Fig. 4. The original and corrected log K_{OA} data of aromatic pollutants studied.

From Fig. 4, it can be seen that the deviations between the experimental and predicted log K_{OA} values were significantly reduced after the correction. Therefore, the dimer has a great effect on the estimation of log K_{OA} values, and this should be taken into account when log K_{OA} values are theoretically predicted. Based on the QSAR model for the dimer percentage, dimer percentages of some PAHs, PBDEs, PCNs and PCDDs without experimental log K_{OA} values were predicted, the results of which are listed in Tables S5, S6, S7 and S8, respectively. Subsequently, the predicted log K_{OA} values of these aromatic pollutants were corrected for the dimer effect following Eq. (5). The corrected ΔG and log K_{OA} values are listed in Table S9, which could provide basic data for these aromatic pollutants without experimental log K_{OA} values.

3.4. Comparison with previous studies

Parnis et al. (2015) estimated log K_{OA} values of some PAHs and their nitrogen-, oxygen- and sulfur-containing derivatives by using the COSMO-RS method and simple linear regression (SLR). Papa et al. (2009) developed a QSPR model to predict log K_{OA} values of PBDEs according to the OECD principles for regulatory acceptability of QSARs with ordinary least squares (OLS). Wang et al. (2017) constructed 3D-QSAR models for K_{OA} values of PCNs by using the comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) techniques based on 32 experimental values of PCNs. Vikas and Chayawan (2015) developed a QSPR model to estimate log K_{OA} values of PCDD/Fs based on a single descriptor, the electron-correlation contribution to the correlation-energy, with the SLR method. The statistical parameters of these models are listed in Table 2.

From Table 2, it can be seen that the statistics of R^2, *RMSE* and Q^2_{CV} of this study were slightly inferior to previously reported results (Papa et al., 2009; Parnis et al., 2015; Vikas and Chayawan, 2015; Wang et al., 2017). However, more aromatic pollutants from four kinds of aromatic pollutants were included. It is known that the performance of a predictive

Table 2
Statistical parameters of predictive models for log K_{OA} of selected aromatic pollutants.

Model	Aromatic Pollutants	Statistical technique	N_{TRA}/N_{VAL}	R^2	*RMSE*	Q^2_{CV}	Reference
COSMO-RS	PAHs	SLR	29	0.991	1.640		Parnis et al. (2015)
QSPR	PBDEs	OLS	24/6	0.961	0.280	0.950	Papa et al. (2009)
QSAR	PCNs	CoMFA	32/11	0.992	0.145	0.966	Wang et al. (2017)
QSAR	PCNs	CoMSIA	32/11	0.991	0.145	0.976	Wang et al. (2017)
QSPR	PCDDs	SLR	10	0.983	0.250	0.959	Vikas and Chayawan (2015)
ΔG_{OA}	PAHs, PBDEs, PCNs, PCDDs	SLR	80	0.935	0.452	0.928	This study

model is not only dependent on these statistical parameters, but also on other characteristics, such as the application domain and the dependency on experimental data. Compared with other studies, the ΔG_{OA} model developed in this study covers more kinds of aromatic pollutants, has a more universal application domain, and is less dependent on experimental data. Furthermore, the dimer effect on the prediction of K_{OA} was systemically and explicitly analyzed for aromatic pollutants, rather than implied in the regression coefficient of statistical analyses. Therefore, the ΔG_{OA} model developed in this study could be a promising alternative method for estimating log K_{OA} values of aromatic pollutants.

4. Conclusions

In the present study, generalized predictive models for log K_{OA} values of aromatic pollutants were developed based on ΔG_{OA} and corrected by the dimer effect. It is found that dimers, driven by π-π interactions, could greatly affect the apparent K_{OA} of aromatic pollutants. London dispersion and exchange-repulsion terms were identified to be dominant components of the underlying π-π interactions. A universal QSAR model was developed to estimate the dimer percentages based on five molecular structural descriptors relevant to the π-π interactions. After calibration with the dimer effect, estimations of log K_{OA} were consistent with experimental values. Therefore, the solvation free energy model could be an alternative method for estimating log K_{OA} values and the dimer effect should be taken into consideration when investigating the partition behavior of aromatic pollutants.

Credit authors statement

Wanran Li: Conceptualization, Methodology, Validation, Formal analysis, Writing - original draft. Dezhi Chen: Investigation, Validation. Shuhua Chen: Conceptualization, Methodology. Jing Zhang: Conceptualization, Methodology. Guobin Song: Investigation, Writing - review & editing. Yawei Shi: Formal analysis, Writing - review & editing. Ya Sun: Investigation. Guanghui Ding: Conceptualization, Methodology, Formal analysis, Validation, Writing - review & editing. Willie J.G.M. Peijnenburg: Conceptualization, Methodology, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work is financially supported by the National Natural Science Foundation of China (51479016 and 42177267), the Science and Technology Innovation Foundation of Dalian (2019J13FZ128), and the Postgraduate Innovation Project of Dalian Maritime University (BSCXXM026).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chemosphere.2022.136608.

References

Abdalla, S., Fink, R.F., 2016. Analyzing interaction energy of polycyclic aromatic hydrocarcons (PAHs) dimer. In: Ramasami, P., Gupta Bhowon, M., Jhaumeer Laulloo, S., Li Kam Wah, H. (Eds.), Crystallizing Ideas-The Role of Chemistry. Springer, Cham, pp. 113–126.
Arunan, E., Gutowsky, H.S., 1993. The rotation spectrum, structure and dynamics of a benzene dimer. J. Chem. Phys. 98, 4294–4296.
Atkinson, E., Crocker, D.E., Ortiz, R.M., 2018. Endocrine Systems. Encyclopedia of Marine Mammals, third ed. Elsevier, Amsterdam, pp. 318–328.
Boys, S., Bernardi, F., 1970. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. Mol. Phys. 19, 553–566.
Byrd, E.F.C., Rice, B.M., 2006. Improved prediction of heats of formation of energetic materials using quantum mechanical calculations. J. Phys. Chem. A 110, 1005–1013.
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.
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.
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.
Grimme, S., Antony, J., Enrlich, S., Krieg, H., 2010. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. J. Chem. Phys. 132, 154104.
Grimme, S., Enrlich, S., Goerigk, L., 2011. Effect of the damping function in dispersion corrected density functional theory. J. Comput. Chem. 32, 1456–1465.
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., 1998. Octanol-air partition coefficient for describing particle/gas partitioning of aromatic compounds in urban air. Environ. Sci. Technol. 32, 1494–1502.
Harner, T., Bidleman, T.F., 1998. Measurement of octanol-air partition coefficients for polycyclic aromatic hydrocarbons and polychlorinated naphthalenes. J. Chem. Eng. Data 43, 40–46.
Harner, T., Green, N.J.L., Jones, K.C., 2000. Measurements of octanol-air partition coefficients for PCDD/Fs: a tool in assessing air-soil equilibrium status. Environ. Sci. Technol. 34, 3109–3114.
Harner, T., Shoeib, M., 2002. Measurements of octanol-air partition coefficients (KOA) for polybrominated diphenyl ethers (PBDEs): predicting partitioning in the environment. J. Chem. Eng. Data 47, 228–232.
Humphrey, W., Dalke, A., Schulten, K., 1996. Vmd – visual molecular dynamics. J. Mol. Graph. 14, 33–38.
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.
Jiao, Y.C., Liu, Y., Zhao, W.J., Wang, Z.X., Ding, X.L., Liu, H.X., Lu, T., 2017. Theoretical study on the interactions of halogen-bonds and pnicogen-bonds in phosphine derivatives with Br2, BrCl, and BrF. Int. J. Quant. Chem. 117, e25443.
Jin, X.C., Fu, Z.Q., Li, X.H., Chen, J.W., 2017. Development of polyparameter linear free energy relationships models for octanol-air partition coefficients (KOA) of diverse chemicals. Environ. Sci.: Process. Impacts 19, 300–306.
Johnson, E.R., Keinan, S., Mori-Sánchez, P., Contreras-García, J., Cohen, A.J., Yang, W. T., 2010. Revealing noncovalent interactions. J. Am. Chem. Soc. 132, 6498–6506.
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, first ed. Elsevier Science, Ltd., Amsterdam, the Netherlands.
Li, W.R., Ding, G.H., Gao, H., Zhuang, Y.T., Gu, X.Y., Peijnenburg, W.J.G.M., 2020. Prediction of octanol-air partition coefficients for PCBs at different ambient temperatures based on the solvation free energy and the dimer ratio. Chemosphere 242, 125246.
Li, W.R., Fan, W.C., Ding, G.H., Zhang, J., Chen, S.H., Shi, Y.W., Sun, Y., Peijnenburg, W. J.G.M., 2022. Aggregation Behavior of Typical Aromatic Pollutants in the Octanol Phase and its Influence on the Octanol-Air Partition Coefficient. Unpublished results.
Liu, Z.Y., Lu, T., Chen, Q.X., 2021. Intermolecular interaction characteristics of the all-carboatomic ring, cyclo[18]carbon: forcusing on molecular adsorption and stacking. Carbon 171, 514–523.
Lu, T., Chen, F.W., 2012a. Quantitative analysis of molecular surface based on improved Marching Tetrahedra algorithm. J. Mol. Graph. Model. 38, 314–323.
Lu, T., Chen, F.W., 2012b. Multiwfn: a multifunctional wavefunction analyzer. J. Comput. Chem. 33, 580–592.
Lu, T., Chen, Q.X., 2020. van der Waals potential: an important complement to molecular electrostatic potential in studying intermolecular interactions. J. Mol. Model. 26, 315.
Lu, T., Manzetti, S., 2014. Wavefunction and reactivity study of benzo[a]pyrene diol epoxide and its enantiomeric forms. Struct. Chem. 25, 1521–1533.
Mackay, D., Callcott, D., 1998. Partitioning and physical chemical properties of PAHs. In: Neilson, A.H. (Ed.), PAHs and Related Compounds. Spring, Berlin, German, pp. 325–345.
Manzetti, S., Lu, T., 2012. The geometry and electronic structure of aristolochic acid: possible implications for a frozen resonance. J. Phys. Org. Chem. 26, 473–483.
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. Theor. 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.
Murray, J.S., Politzer, P., 2011. The electrostatic potential: an overview. WIREs Comput. Mol. Sci. 1, 153–163.
Nedyalkova, M., Madurga, S., Tobiszewski, M., Simeonov, V., 2019. Calculating the partition coefficients of organic solvent in octanol/water and octanol/air. J. Chem. Inf. Model. 59, 2257–2263.
Odabasi, M., Cetin, E., Sofuoglu, A., 2006. Determination of octanol-air partition coefficients and supercooled liquid vapor pressures of PAHs as a function of temperature: application to gas-particle partitioning in an urban atmosphere. Atoms. Environ. 40, 6615–6625.
Papa, E., Kovarich, S., Gramatica, P., 2009. Development, validation and inspection of the applicability domain of QSPR models for physicochemical properties of polybrominated diphenyl ethers. QSAR Comb. Sci. 28, 790–796.
Parnis, J.M., Mackay, D., Harner, T., 2015. Temperature dependence of Henry's law constants and KOA for simple and heteroatom-substituted PAHs by COSMO-RS. Atoms. Environ. Times 110, 27–35.
Podeszwa, R., Szalewicz, K., 2008. Physical origins of interactions in dimers of polycyclic aromatic hydrocarbons. Phys. Chem. Chem. Phys. 10, 2735–2746.
Politzer, P., Murray, J.S., Peralta-Inga, Z., 2001. Molecular surface electrostatic potential in relation to noncovalent interactions in biological systems. Int. J. Quant. Chem. 85, 676–684.
Sato, T., Tsuneda, T., Hirao, K., 2005. A density-functional study on π-aromatic interaction: benzene dimer and naphthalene dimer. J. Chem. Phys. 123, 104307.
Sherrill, C.D., 2013. Energy component analysis of π interactions. Acc. Chem. Res. 46, 1020–1028.
Sinnokrot, M.O., Sherrill, C.D., 2004. Highly accurate coupled cluster potential energy curves for the benzene dimer: sandwich, T-shaped, and parallel-displaced configurations. J. Phys. Chem. A 108, 10200–10207.
Szalewicz, K., 2012. Symmetry-adapted perturbation theory of intermolecular forces. WIREs Comput. Mol. Sci. 2, 254–272.
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.
Treves, K., Shragina, L., Rudich, Y., 2001. Measurement of octanol-air partition coefficients using solid-phase microextraction (SPME) - application to hydroxy alkyl nitrates. Atoms. Environ. 35, 5843–5854.
Tsuzuki, S., Honda, K., Uchimaru, T., Mikami, M., Tanabe, K., 2002. Origin of attraction and directionality of the π/π interaction: model chemistry calculations of benzene dimer interaction. J. Am. Chem. Soc. 124, 104–122.
Vikas, Chayawan, 2015. Single-descriptor based quantum-chemical QSPRs for physico-chemical properties of polychlorinated-dibenzo-p-dioxins and -dibenzo-furans (PCDD/Fs): exploring the role of electron-correlation. Chemosphere 118, 239–245.
Viluksela, M., Pohjanvirta, R., 2019. Multigenerational and transgenerational effects of dioxins. Int. J. Mol. Sci. 20, 2947.
Wang, X.L., Gu, W.E., Guo, E.M., Cui, C.Y., Li, Y., 2017. Assessment of long-range transport potential of polychlorinated naphthalenes based on three-dimensional QSAR models. Environ. Sci. Pollut. Res. 24, 14802–14818.
Wang, X.P., Wang, C.F., Zhu, T.T., Gong, P., Fu, J.J., Cong, Z.Y., 2019. Persistent organic pollutants in the polar regions and the Tibetan Plateau: a review of current knowledge and future prospects. Environ. Pollut. 248, 191–208.
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.
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.
Wold, S., Sjöström, M., Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. Chemometr. Intell. Lab. 58, 109–130.
Xu, S.H., Kozerski, G., Mackay, D., 2014. Critical review and interpretation of environmental data for volatile methylsiloxanes: partition properties. Environ. Sci. Technol. 48, 11748–11759.
Zeinalipour-Yazdi, C.D., Pullman, D.P., 2006. Correlation of polarizabilities with van der Waals interactions in π-systems. J. Phys. Chem. B 110, 24260–24265.
Zhang, C.Y., 2011. Shape and size effects in π-π interactions: face-to-face dimers. J. Comput. Chem. 32, 152–160.