

RESEARCH ARTICLE

Quantitative structure activity relationship (QSAR) modeling for adsorption of organic compounds by activated carbon based on Freundlich adsorption isotherm

Siyi Ding¹, Yefan Yang², Ting Tan^{3,4}, Qunshan Wei^{3,4*}, Zhemin Shen⁵, Qiong Liu^{3,4}, Xinshan Song^{3,4}, Yuhui Wang^{3,4}, Charles Nzila⁶, Christopher W.K. Chow⁷

1 Institute of Artificial Intelligence, Donghua University, Shanghai, China, **2** College of Mechanical Engineering, Donghua University, Shanghai, China, **3** College of Environmental Science and Engineering, Donghua University, Shanghai, China, **4** State Environmental Protection Engineering Center for Pollution Treatment and Control in Textile Industry, Donghua University, Shanghai, China, **5** School of Environmental Science and Engineering, Shanghai Jiao Tong University, Shanghai, China, **6** School of Engineering, Moi University, Eldoret, Kenya, **7** Sustainable Infrastructure and Resource Management (SIRM), UniSA STEM, University of South Australia, Mawson Lakes, South Australia, Australia

* qswei@dhu.edu.cn



OPEN ACCESS

Citation: Ding S, Yang Y, Tan T, Wei Q, Shen Z, Liu Q, et al. (2025) Quantitative structure activity relationship (QSAR) modeling for adsorption of organic compounds by activated carbon based on Freundlich adsorption isotherm. PLoS One 20(12): e0338483. <https://doi.org/10.1371/journal.pone.0338483>

Editor: Alaa El Din Mahmoud, Faculty of Science, Alexandria University, EGYPT

Received: March 13, 2025

Accepted: November 24, 2025

Published: December 15, 2025

Copyright: © 2025 Ding et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: All relevant data are within the manuscript and its [Supporting Information](#) files.

Funding: This study was financially supported by the AI-Enhanced Research Program of Shanghai Municipal Education Commission

Abstract

The Freundlich isotherm parameters K and $1/n$ are typically regarded as empirical constants. However, the underlying theoretical basis for the widespread applicability of the Freundlich isotherm in describing adsorption processes for diverse organic compounds remains unclear. In this study, we successfully elucidated the reason by developing two optimal quantitative structure-activity relationship (QSAR) models: one correlating K with quantum chemical parameters and another linking $1/n$ to these parameters. The modeling results demonstrated that both K and $1/n$ exhibit strong correlations with specific quantum chemical descriptors, indicating that the empirical Freundlich isotherm's applicability is fundamentally linked to the molecular structural characteristics of organic compounds. Key quantum parameters influencing K were identified as $\sum q(O+N)$, $q(CH^+)_{max}$, ELUMO, Fukui $(-)$ $_{max}$, and Wiberg $(C-C)_{min}$, suggesting that charge distribution, carbon bond energy, and active site energy are the primary factors governing adsorption efficiency on activated carbon. The QSAR model for $1/n$ yielded similarly novel and consistent insights, showing that the value of $1/n$ also correlated with molecular structural characteristics. Both models were rigorously validated and confirmed to be stable, robust, and accurate through standard statistical evaluations. These QSAR models can now be employed to identify whether an organic compound would conform to the Freundlich Isotherm and predict the adsorption efficiency of this compound by activated carbon based on their quantum chemical parameters. As to the practical implications, this study provides a convenient reference method for assessing the applicability of activated carbon adsorption in treating emerging organic pollutants in drinking water plants and a

in the form of a project award (SMEC-AI-DHUZ-07) received by SD, YY, TT, and QW. This study was also financially supported by the National Key Research and Development Program of China in the form of an award (2023YFC3707702) received by ZS.

Competing interests: 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.

theoretical foundation for developing intelligent management systems in water treatment facilities.

1 Introduction

In recent years, the widespread use of organic compounds (e.g., halogenated hydrocarbons, benzene, and pollutants containing nitrogen or oxygen functional groups) in industrial and agricultural production has led to their entry into the water environment through the pathways of agricultural runoff and industrial wastewater, which has brought about great risks and impacts on the environment and human health [1,2]. Therefore, it is necessary to effectively control the impact of these organic compounds on drinking water. Activated carbon adsorption (ACA) technology remains employed in drinking water treatment plants worldwide due to its high efficiency, easy operation and stable performance in removing some organic pollutants. However, for some emerging organic compounds, the removal efficiencies by ACA are still not clear, which needs experimental assessment individually. Nevertheless, the experimental assessment of the removal efficiencies for specific pollutants constitutes a labor-intensive process, and most of small or even middle-size drinking water plants don't have the sufficient analytic instruments and technicians. Therefore, establishing a quantitative structure-activity relationship (QSAR) model based on quantum chemical parameters of organic molecular structures holds significant reference value for current drinking water treatment plants to evaluate their activated carbon units in controlling emerging organic contaminants.

As a well-established cheminformatics methodology, QSAR quantitatively relates measurable biological/chemical activities of compounds to their structural parameters using mathematical models and molecular descriptors [3,4]. Typically, it can reliably predict key compound properties including metabolic activity and toxicity profiles using only molecular descriptors derived from compound structures [5]. Current adsorption models predominantly focus on hydrophobic interactions as the interpretation of adsorption results, particularly for nonpolar or weakly polar organic compounds [6]. Hobbs et al. [7] employed the Freundlich isotherm to develop a QSAR model for aromatic compound adsorption on powdered activated carbon. However, the structural homogeneity of the selected adsorbates limited the universal applicability of the model. Blum et al. [8] utilized molecular connectivity and $(X/C)_{\min}$ to develop a QSAR model in order to assess the adsorption of organic compounds (both aromatic and aliphatic) by activated carbon. However, molecular structure quantum parameters were not taken into account, therefore the QSAR model could not fully predict or explain the adsorption properties of these compounds. Rao et al. [9] built three different QSAR models: the octanol-water partition coefficient model (K_{ow} model), the linear solvation energy model (LSE model), and the molecular connectivity indices theory (MCI) model, which were only validated by the experimental adsorption data of Dewulf et al. [10] and Mader et al. [11] and were not correlated with quantum chemical parameters of molecular structures, limiting the model's applicability. Utilizing partial least squares regression (PLS), Kamlet et al. established

multivariate linear regression equations to elucidate specific solute-adsorbent interaction mechanisms. The models integrated three key solvation parameters – dipolarity/polarizability (π^*), hydrogen-bond donor acidity (α), and hydrogen-bond acceptor basicity (β) – quantified through systematic analysis of removal kinetics of mono-nitroaromatic by activated carbon adsorption [12]. However, the structures of the compounds studied are relatively homogeneous, and therefore the adsorption mechanism may not be suitable for other kinds of organics with different structures.

According to the summary of existing results shown above, it can be seen that there were few researches establishing QSAR models for the key parameters (K and $1/n$) of Freundlich Isotherm, which are commonly considered as empirical constants determining the removal efficiencies of organics by activated carbon adsorption. It is still not very clear why the empirical Isotherm model has been applicable to so many organics for many years.

Therefore, the first aim of this study is to reveal whether the key parameters (K and $1/n$) of Freundlich Isotherm are correlated with some of the quantum parameters of organic molecular structure, then to explain why the empirical Isotherm model can be applicable for many organic compounds. Is this really a coincidence, or is it a special coincidence that just happens to fall on the track of the laws of the world? If the QSAR model can be constructed successfully, it would be a significant innovation result to let the question get an exact answer. The findings will further reveal which molecular quantum parameters determine the applicability of the Freundlich isotherm, thereby provide novel theoretical insights into this classical adsorption model.

The second aim of this study is to provide a kind of useful calculation model for current drinking water plants to evaluate the possible efficiencies of their activated carbon units when the plants meet various emerging organic compounds. The results would help the plants to overcome their weak points on the analytical instruments and technicians.

In future intelligent water treatment plant management systems, core processes like activated carbon adsorption will require fundamental chemical interaction models and performance evaluation frameworks. This study will develop QSAR models for Freundlich parameters (K and $1/n$) of organic compound adsorption on activated carbon, which will significantly contribute to building next-generation smart water management systems.

2 Materials and methods

2.1 Data sets

Table 1 summarizes the Freundlich adsorption isotherm (Eq. (1)) K and $1/n$ for these 47 organic compounds, which were extracted from the Contaminant Control List of Drinking Water developed by the U.S. Environmental Protection Agency and related literature [5,8,13–16]. These compounds structurally include halogenated hydrocarbons (especially chlorinated alkanes and aromatic hydrocarbons), benzene rings and their derivatives, and nitrogen or oxygen containing functional groups. All data were obtained from adsorption experiments using Fitratorb-400 coal activated carbon. The concentrations for all organic compounds during the adsorption experiment were in the same measurable range of the analytical method (normally referred to USA EPA method). Uniform criteria were followed in screening the data: (1) the experimental temperatures were controlled within the range of room temperatures (25°C); (2) the final state data that reached adsorption equilibrium were chosen (contact time was greater than 48 h). The above standardized screening ensured the maximum baseline comparability of the data.

The literature data compiled in this study were obtained from batch equilibrium experiments conducted at concentrations within the measurable range of the analytical methods for most compounds. The pH values of the solutions varied from 5.3 to 8.0 in the data extracted from literature because the adsorption experiments were performed without pH control, in which the organic compounds were dissolved into the pure water directly and kept in their natural states. In addition, the K with excessive variance data was excluded considering the applicability of the data and the stability of the later model. To ensure the robustness of the developed model, we followed this principle in organizing the extracted data from literature that the data conforms to a Normal Distribution (i.e., Gaussian Distribution) were retained but the data deviating from the overall Normal Distribution are excluded.

Table 1. Fitting parameters of Freundlich adsorption isotherms for 47 organic compounds.

Name	K	1/n	Name	K	1/n	Name	K	1/n
Benzene	1260	0.533	2,4-Dinitrotoluene	96100	0.157	Diquat	2260	0.242
Toluene	5010	0.429	1,3-Dichloropropane	897	0.497	Dinoseb	30400	0.279
o-Chlorotoluene	23200	0.378	1,1-Dichloropropene	2670	0.374	Picloram	23400	0.18
o-Dichlorobenzene	19300	0.378	1,1,1,2-Tetrachloroethane	1070	0.604	Metribuzin	48700	0.193
Ethyl benzene	9270	0.415	1,2-Dibromoethane	888	0.471	Aldicarb	8270	0.402
Styrene	12200	0.479	1,3,5-Trichlorobenzene	63800	0.324	Oxamyl	1740	0.793
p-Xylene	12600	0.418	Dibromochloromethane	585	0.636	Lindane	15000	0.433
p-Chlorotoluene	35900	0.34	trans-1,2-Dichloroethylene	618	0.452	Alachlor	81700	0.257
Bromobenzene	17200	0.364	cis-1,2-Dichloroethylene	202	0.587	Atrazine	38700	0.291
Chlorobenzene	9170	0.348	Bromoform	929	0.665	Carbofuran	16400	0.408
1,2-Dichloroethane	129	0.533	1,1,1-Trichloroethane	335	0.531	Dicamba	33100	0.147
Isophorone	9750	0.271	Bromodichloromethane	241	0.655	Glyphosate	87600	0.119
1,2-Dichloropropane	313	0.597	1,1-Dichloroethene	470	0.515	Metolachlor	98200	0.125
1,2,3-Trichloropropane	1080	0.613	1,1,2-Trichloroethane	365	0.652	Simazine	31300	0.227
tert-Butyl methyl ether	218	0.479	Trichloroethylene	2000	0.482	Cyanazine	102000	0.126
2,4,5-trichlorophenoxy acetic acid	43000	0.21	Dibromochloropropane	6910	0.501	—	—	—

<https://doi.org/10.1371/journal.pone.0338483.t001>

$$Q_e = KC_e^{1/n} \quad (1)$$

where Q_e is the adsorbed amount at equilibrium, mg/g; C_e is the concentration at equilibrium, mg/L; K is Freundlich's constant; 1/n is the adsorption index. The concentrations for all organic compounds during the adsorption experiment were in the same measurable range of the analytical method (normally referred to USA EPA method).

2.2 Calculation method of structural parameters

Studies have shown that chemical structural parameters are closely related to the properties of compounds [17], and thus focuses on common parameters that represent the structural properties of organic compounds, which are shown in S1 Table. These parameters were drawn and calculated using ChemDraw, Gaussian, Material Studio, and Multiwfn [18] software. All calculations were performed using the Density Functional Theory Method (DFT) [4,19]. The DFT calculations of compounds are performed in a consistent identical and theoretical state. The core principle of Density Functional Theory (DFT) calculations lies in the assumption that, under a unified idealized state, the electronic structure of a system (such as electron density, energy levels, and orbital energies) is determined by solving the Schrödinger equation, thereby reflecting the intrinsic properties of the material. This methodology is applicable to diverse systems, including vacuum, solution, and surface environments. The B3LYP method and 6 - 311G (d,p) basis set were chosen to obtain the structural parameters of the substance by Gaussian 09 [20]. The DMol3 module and the GGA-BLYP method were used to calculate the energies of the optimized structures by Material Studio 7.0 [21]. The details of the molecular parameters of each organic compound are shown in S2 Table. It should be noted that all quantum chemical calculations were performed under vacuum conditions. While this approach is standard in QSAR studies for computational efficiency and has been shown to yield robust correlation trends for adsorption affinity, it does not explicitly account for aqueous solvation effects. Future investigations will employ implicit solvation models (e.g., SMD, PCM) and descriptors like solvent-accessible surface area (SASA) to better simulate the aqueous-phase adsorption environment and refine the predictions.

2.3 Model validation methods

In this study, QSARINS software was used to build the model using the multiple linear regression method [21,22], and the model was validated by Ordinary Least Squares (OLS) method [23] with Y-randomization and Leave-Many-Out cross-validation method. The quality of the constructed QSAR was evaluated by different evaluation metrics to test its robustness and predictive ability [24]. The information of evaluation indexes is shown in S2 Table.

Y-randomization validation [25,26] is mainly to determine the stability of the model by testing the chance relationship between the dependent and independent variables. This method disrupts the dependent variables firstly and then builds a new model for the randomly disrupted dependent variable and initial independent variable. The process is repeated several times. For a stable QSAR model, R^2_{yrand} and Q^2_{yrand} should be lower compared to the original model's R^2 and Q^2 .

Leave-Many-Out cross-validation [27,28] was used to study the behavior of the model when many compounds were excluded. Multiple samples are drawn from the data set at a time, and the remaining samples are used to build a model. The drawn samples are predicted by the constructed model. This above process is repeated several times.

2.4 Application domains

In this study, a Williams plot [25,29] was applied to visualize the applicable domain. The accuracy of the optimal QSAR model prediction was judged based on the standardized residuals (σ) and leverage (h_i). The standardized residuals are defined in the following equation (2):

$$\sigma = \frac{(y_i - \hat{y}_i)}{\sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-1}}} \quad (2)$$

where y_i is the experimental value of the compound, \hat{y}_i is the predicted value the compound, n is the number of compounds in the training set. The ideal range of σ is $(-3, 3)$. The leverage is defined in the following equation (3):

$$h_i = x_i (X^T X)^{-1} x_i^T (i = 1, 2, \dots, n) \quad (3)$$

where x_i is a vector of descriptor for the compound, X is the descriptor matrix in the training set. The alert leverage value (h^*) is defined in the following equation (4):

$$h^* = \frac{3(m+1)}{n} \quad (4)$$

where m is the number of predictor variables and n is the number of compounds in the training set. It was only $h_i < h^*$ that it was considered compliant.

3 Results and analysis

3.1 Experimental results of organic compound

In the Freundlich isotherm, the larger the value of K , the higher the adsorption capacity of the adsorbent at a given pressure; and the larger the value of n , the stronger the nonlinearity of the adsorption process. Both determine the shape of the Freundlich isotherm curve and therefore together determine the characteristics of the adsorption process and the performance of the adsorbent. The K and $1/n$ values of the 47 compounds summarized are shown in Table 1. The K and $1/n$ values varied from compound to compound. Among the 47 compounds selected, glyphosate had the smallest Freundlich adsorption index $1/n$ (0.119), corresponding to K of 87,600, and oxamyl had the largest adsorption index $1/n$ (0.793), corresponding to K of 1740. The difference in the Freundlich adsorption index $1/n$ of the above two compounds is about

0.674, and the difference in the K value is about 85860. It can be seen that the structural differences of these organic compounds are significant, resulting in a wide range of adsorption properties among them. Therefore, the model constructed on the basis of the above compounds may have wider applicability.

3.2 Calculation results of the structure of compounds

The structural parameters of the 47 organic compounds studied are shown in [S2 Table](#). The differences in dipole moments among these 47 organic compounds indicate that the polarities of the organic compounds vary considerably. The total molecular energy (E) represents the interaction of kinetic and potential energy between electrons. Its maximum value is -232.333 (benzene), and the minimum value is -7761.158 (bromoform). There is a 33-fold difference between these two values, which indicates the diversity between the structures of the studied organic compounds.

$q(\text{CH}^+)_{\text{max}}$ has a maximum of 0.198 e (dicamba) and a minimum value of 0.039 e (tert-butyl methyl ether); $q(\text{CH}^+)_{\text{min}}$ has a maximum of 0.129 e (picloram) and a minimum of 0.020 e (metolachlor); $q(\text{C}^-)_{\text{max}}$ has a maximum of 0.219 e (glyphosate) and a minimum of -0.045 e (benzene); $q(\text{C}^-)_{\text{min}}$ has a maximum of 0.034 e (bromoform) and a minimum of -0.137 e (1,2-dichloropropane). $\sum q(\text{O} + \text{N})$ has a maximum of 0.154 e (diquat) and a minimum of -1.554 e (glyphosate), and 29 compounds did not contain oxygen and nitrogen. $\sum q(-)/N_c$ has a maximum of 0.060 e (diquat) and a minimum of -0.061 e (1,2-dichloroethane). These results demonstrate the charge distribution of the organic compounds studied, which is important for analyzing the adsorption of organic compounds by activated carbon.

The Wiberg index [30] is used to quantify the degree of bonding between atoms. This parameter can be used to analyze the bonding properties of molecules and to predict the breaking and formation of chemical bonds [31,32]. $\text{Wiberg}(\text{C}-\text{C})_{\text{max}}$ and $\text{Wiberg}(\text{C}-\text{C})_{\text{min}}$ represent the bonding strengths of the carbon-carbon bond in a molecule. In this study, $\text{Wiberg}(\text{C}-\text{C})_{\text{max}}$ has a maximum of 4.083 (1,1,1,2-tetrachloroethane) and a minimum of 3.893 (1,2-dichloroethane). The maximum of $\text{Wiberg}(\text{C}-\text{C})_{\text{min}}$ is 3.963 (benzene) and the minimum is 1.072 (1,3-dichloropropane).

E_{HOMO} and E_{LUMO} represent the charge transfer that occurs within the molecules of organic compounds [33]. E_{LUMO} indicates the ability of organic compounds to accept electrons, while the opposite is true for E_{HOMO} . The compound with the maximum E_{LUMO} value is tert-butyl methyl (0.040 eV), and the compound with the minimum value is diquat (-0.145 eV), which indicates that diquat is more likely to be able to obtain electrons compared to other organic compounds. The Fukui indices (Fukui) are key to describing the order of decomposition of the molecular structure of organic compounds [19,34]. It includes the electrophilic Fukui indices, the nucleophilic Fukui indices, and the pro-radical Fukui indices [35]. Among them, $\text{Fukui}(-)_{\text{max}}$ has a maximum of 0.471 e (bromodichloromethane) and a minimum of 0.087 e (diquat). The descriptor bond level reflects the stability of the chemical bonds within a compound molecule. The maximum value of Bond orders(C-C)_{min} is -0.658 (1,3-dichloropropane) and the minimum value is -1.61 (dinoseb). Combined with the correlation plot [Fig 1](#), it can be seen that Bond orders(C-C)_{min}, $\text{Fukui}(-)_{\text{max}}$, E_{HUMO} have a significant correlation with the adsorption effect, which indicates that bond energy, active site, charge distribution, and orbital energy have a strong influence on the adsorption effect of activated carbon.

3.3 Correlation analysis of experimental results with structural parameters

[Fig 1](#) shows the correlation between K, 1/n and 31 parameters. It can be seen that there is consistency in the correlation between these two response values (K, 1/n) and the structural parameters, which also indirectly indicates that there is a certain correlation relationship between K and 1/n in the Freundlich adsorption isotherm. According to the absolute value of the correlation coefficient, it can be seen that Bond orders(C-C)_{min} (0.694), $\text{Fukui}(-)_{\text{max}}$ (0.611), $\sum q(\text{O} + \text{N})$ (0.532), and $\text{Fukui}(-)_{\text{min}}$ (0.515) showed a significant positive correlation with the response values 1/n, while $\sum q(\text{H})$ (-0.598), E_{HUMO} (-0.463), $q(\text{C}^-)_{\text{max}}$ (-0.444), $q(\text{CH}^+)_{\text{max}}$ (-0.414), $\sum q(-)/N_c$ (-0.401) showed significant negative correlation with the response values 1/n. A similar correspondence exists for the response value K. The most significant correlation was found between the response values (K, 1/n) and Bond order(C-C)_{min} with a correlation coefficient of 0.694. This structural

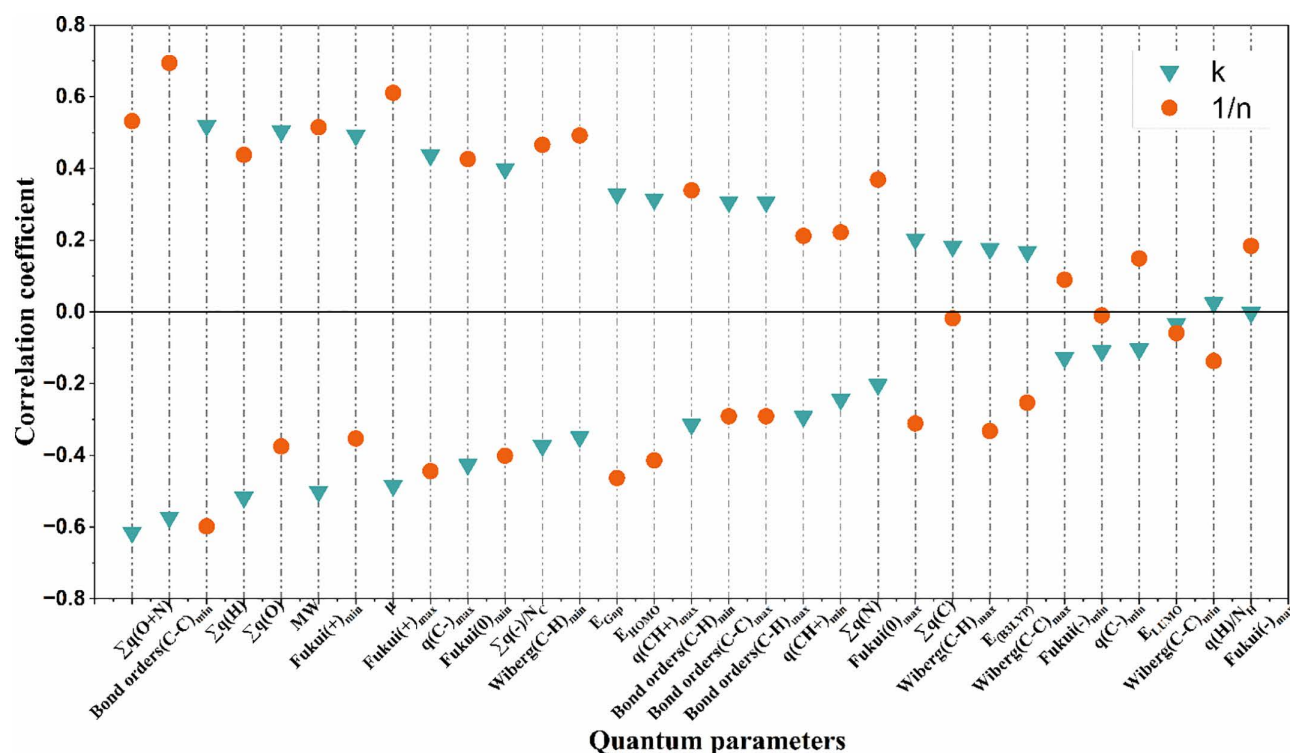


Fig 1. Correlation coefficients of Freundlich parameters (K, 1/n) with 31 molecular parameters.

<https://doi.org/10.1371/journal.pone.0338483.g001>

parameter of Bond order(C-C)_{min} represents the minimum value of the bond orders among the carbon-carbon bonds of a substance, and the magnitude of the value represents the stability of the compound molecule. Another structural parameter is the nucleophilic index (Fukui(-)_{max}), which reveals the possible active sites of the reaction. This result indicates that the response values (K, 1/n) are closely related to the active sites. In conclusion, based on the results of the current correlation analysis, it is clear that Bond orders(C-C)_{min} and Fukui(-)_{max} are the main factors affecting the adsorption of activated carbon.

3.4 QSAR model construction

The 47 organic compounds studied in this paper were randomly divided into a training set of 38 and a test set of 9 in a ratio of 4:1. Among them, 1,2,3-trichloropropane, isophorone, styrene, cis-1,2-dichloroethene, 1,2-dibromoethane, simazine, atrazine, Alachlor, Metolachlor were selected as test sets for the QSAR model with K as the response value, while tert-butyl methyl ether, benzene, bromobenzene, 1,2-dibromoethane, dibromochloroethane, cis-1,2-dichloroethene, and lindane, picloram, Alachlor were selected as test sets for the model with K as the response value. In this study, the optimal model with low multicollinearity and a good correlation with the response values was identified based on the results obtained from the Qsarins software.

Table 2 presents the results of the QSAR models with K as response values using the stepwise method of Qsarins. From Table 2, the gradual introduction of the number of independent variables, the R, R², and R²_{adj} of models 1–5 gradually increase, the SD values gradually become smaller, the F values gradually become larger, and the sig. (significance, see “S3 Table. Meaning of evaluation indexes”, which means the significance values of these five models) are all 0.00. Generally, the determine coefficient R² of the model with a good fit is at least greater than 0.6 [36]. Based on this, models

Table 2. Results of the K-model organized by the Qsarins.

model	Variables of models	R ²	R _{adj} ²	R ² -R _{adj} ²	Sig.	LOF	RMSE _{tr}	MAE _{tr}	RSS _{tr}	SD	F	Q _{LOO} ²	R _{ext} ²	q ²	Q _{ext} ²
1	Σq(O+N)	0.467	0.451	0.016	4*10 ⁻⁶	400.528	20.013	13.329	14419.003	20.59	29.77	0.395	0.345	0.399	0.345
2	Σq(O+N), q(CH ₊) _{max}	0.532	0.503	0.028	4*10 ⁻⁶	351.846	18.757	13.454	12666.441	19.59	18.73	0.426	0.598	0.428	0.598
3	Σq(O+N), q(CH ₊) _{max} , E _{LUMO}	0.684	0.654	0.030	0	237.384	15.407	10.925	8545.805	16.34	20.70	0.553	0.625	0.557	0.625
4	Σq(O+N), q(CH ₊) _{max} , E _{LUMO} , Fukui(-) _{max}	0.728	0.693	0.035	0	266.749	14.305	10.532	7366.427	15.41	23.09	0.601	0.645	0.682	0.645
5	Σq(O+N), q(CH ₊) _{max} , E _{LUMO} , Fukui(-) _{max} , Wiberg(C-C) _{min}	0.801	0.770	0.033	0	148.123	12.171	8.693	5332.415	13.33	24.43	0.699	0.698	0.722	0.699

<https://doi.org/10.1371/journal.pone.0338483.t002>

1 and 2 can be excluded. It is also required that the internal and external validation coefficients q² and Q_{ext}² of the optimal model are both greater than 0.5 [37]. It is clear that models 3, 4, and 5 meet the above requirements, but compared to the rest of the models, R² (0.801) and R_{adj}² (0.770) of model 5 are higher, q² (0.722) and Q_{ext}² (0.699) of model 5 are higher, the standardized estimation error (SD = 13.332) and the root mean square error (RMSE_{tr} = 12.171) of model 5 are minimized. The model 5 has a lower LOF, which indicates that the model has a good fit with these descriptors. The sig. results indicate that the model is statistically significant. The cross-validation coefficients of the internal verification (Q_{LOO}² = 0.699) for internal validation [38], and the external validation coefficient (R_{ext}² = 0.698) for assessing the external prediction ability meet the criterion of greater than 0.5, which indicates that the predictions from the test set are reliable. In addition, the models built from the training set all have small errors for RMSE_{tr} (12.171), MAE_{tr} (8.693), RSS_{tr} (5532.415), which effectively evaluates the external predictive power of the model. Therefore, model 5 was determined to be the optimal model for the K model, which contains five variables, namely Σq(O+N), q(CH₊)_{max}, Fukui(-)_{max}, E_{LUMO}, Wiberg(C-C)_{min}.

Table 3 presents the results of the QSAR models with 1/n as response values. From Table 5, with the gradual introduction of the number of independent variables, the R, R², and R_{adj}² of models 1–7 gradually increase, the SD value gradually becomes smaller, the F value gradually becomes larger, and the sig. The value of these five models is 0.00. According to

Table 3. Results of the 1/n-model organized by the Qsarins.

model	Variables of models	R ²	R _{adj} ²	R ² -R _{adj} ²	Sig.	LOF	RMSE _{tr}	MAE _{tr}	RSS _{tr}	SD	F	Q _{LOO} ²	R _{ext} ²	q ²	Q _{ext} ²
1	Bond orders(C-C) _{min}	0.719	0.711	0.008	4*10 ⁻⁶	0.008	0.087	0.072	0.285	0.089	91.951	0.688	0.001	0.689	0.001
2	Bond orders(C-C) _{min} , Bond orders(C-C) _{max}	0.765	0.751	0.014	1*10 ⁻⁶	—	—	—	—	—	—	—	—	0.731	0.180
3	Bond orders(C-C) _{min} , Bond orders(C-C) _{max} , Σq(-)/N _C	0.810	0.794	0.016	1*10 ⁻⁶	—	—	—	—	—	—	—	—	0.767	0.450
4	Bond orders(C-C) _{min} , Bond orders(C-C) _{max} , Σq(-)/N _C , Wiberg(C-C) _{min}	0.846	0.827	0.019	0	—	—	—	—	—	—	—	—	0.802	0.485
5	Bond orders(C-C) _{min} , Bond orders(C-C) _{max} , Σq(-)/N _C , Wiberg(C-C) _{min} , q(C-) _{max}	0.885	0.867	0.018	0	0.006	0.055	0.047	0.116	0.060	49.406	0.839	0.738	0.840	0.738
6	Bond orders(C-C) _{min} , Bond orders(C-C) _{max} , Σq(-)/N _C , Wiberg(C-C) _{min} , q(C-) _{max} , Fukui(+) _{max}	0.904	0.885	0.019	0.	0.006	0.051	0.043	0.098	0.056	48.567	0.860	0.734	0.861	0.740
7	Bond orders(C-C) _{min} , Bond orders(C-C) _{max} , Σq(-)/N _C , Wiberg(C-C) _{min} , q(C-) _{max} , Fukui(+) _{max} , Fukui(+) _{min}	0.917	0.897	0.019	1*10 ⁻⁶	0.006	0.048	0.039	0.085	0.053	47.166	0.873	0.704	0.873	0.704

<https://doi.org/10.1371/journal.pone.0338483.t003>

the standard requirements of the optimal model, models 1–4 can be excluded. The comparison of the internal and external validation coefficients found that models 5–7 conformed to the criterion of being greater than 0.5. Q_{ext}^2 of model 5 is larger (0.738), and q^2 of model 7 is larger (0.873). But the LOF of model 5 and model 7 are consistent, and other indicators (K_{xx} , MSE_{tr} , MAE_{tr} , and RSS_{tr}) are not showing much difference between the two models. Model 5 has the largest R_{ext}^2 (0.738) and the smallest Q_{LOO}^2 (0.839) in contrast to model 7, which meets the criterion and indicates that the predictions of the test set under model 5 are reliable. Overall, there is not much difference between model 5 and model 7. To avoid overfitting and ensure the prediction performance, model 5 with fewer parameters and simple structure is determined to be the 1/n optimal model following the principle of consistency with the variables of the K model. The model contains Bond orders(C-C)_{min}, Bond orders(C-C)_{max}, $q(\text{C-})_{\text{max}}$, $\sum q(-)/N_{\text{C}}$, Wiberg(C-C)_{min}.

In addition, we can identify potential outliers by the plot of predicted versus experimental response values (Figs 2 and 3). Fig 2 shows the scatter distributions of predicted and experimental values of the K models on the training and test sets. It can be observed that model 5 has a better fit and a denser distribution compared to the other models. Most of the predicted data are distributed around the 1:1 ($y=x$) regression line, only three compounds such as simazine, 2,4,5-trichlorophenoxy acetic acid, and isophorone are outside the range of variation of the residuals. In the scatter plot of the 1/n model, only two compounds (picloram and nematicarb) are slightly further apart from the regression line.

Further analysis showed that the predictive ability of the model built for these structures was limited due to triazines (atrazine and simazine), and carbamate pesticides containing a nitrogen heterocyclic ring (oxamyl), which were underrepresented in the training set, thus leading to a limited predictive ability of the model built for these structures. Therefore, we have already excluded these categories when establishing the K model. The adsorption mechanisms of 2,4,5-trichlorophenoxy acetic acid, isophorone, simazine and picloram may involve hydrophobic interactions, and the

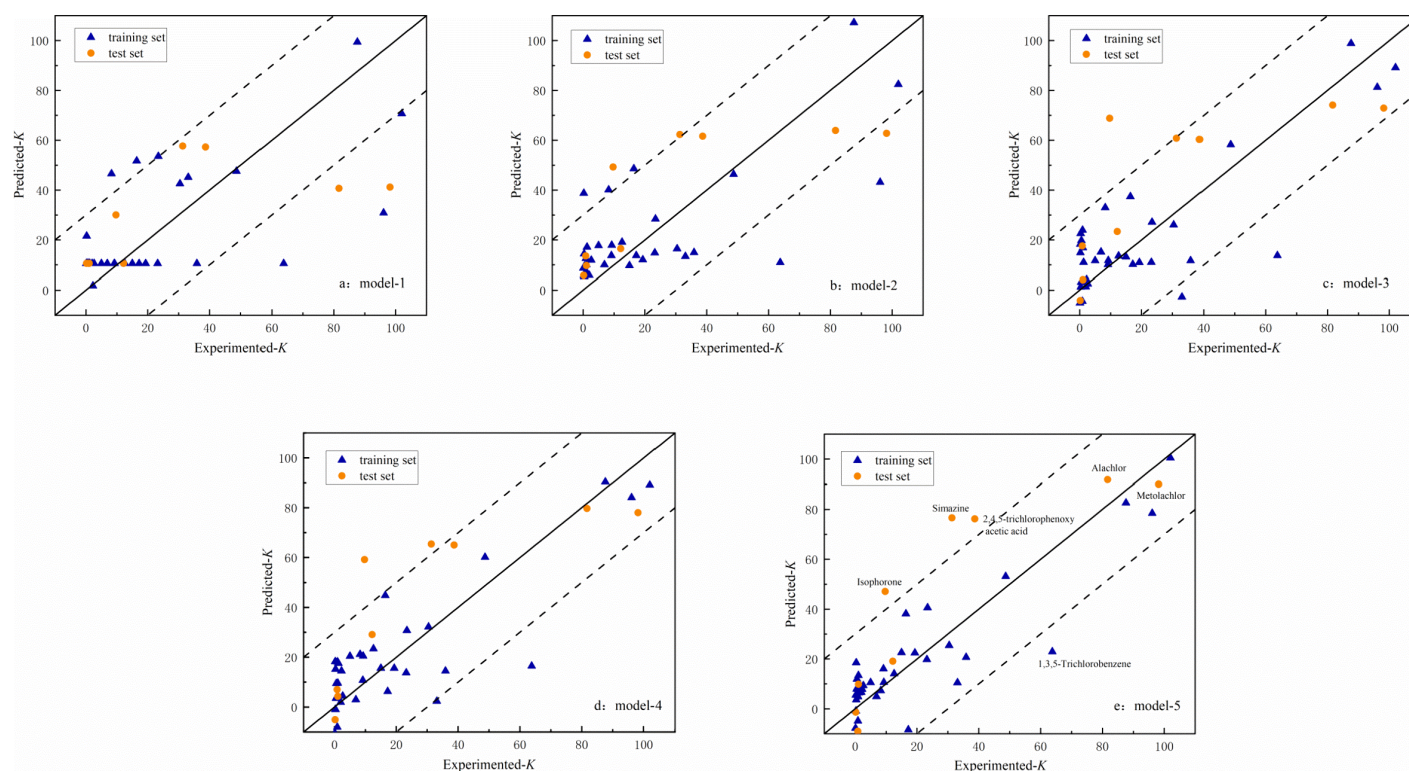


Fig 2. Scatter plot of predicted versus experimental values for models about K; a) model-1; b) model-2; c) model-3; d) model-4; e) model-5.

<https://doi.org/10.1371/journal.pone.0338483.g002>

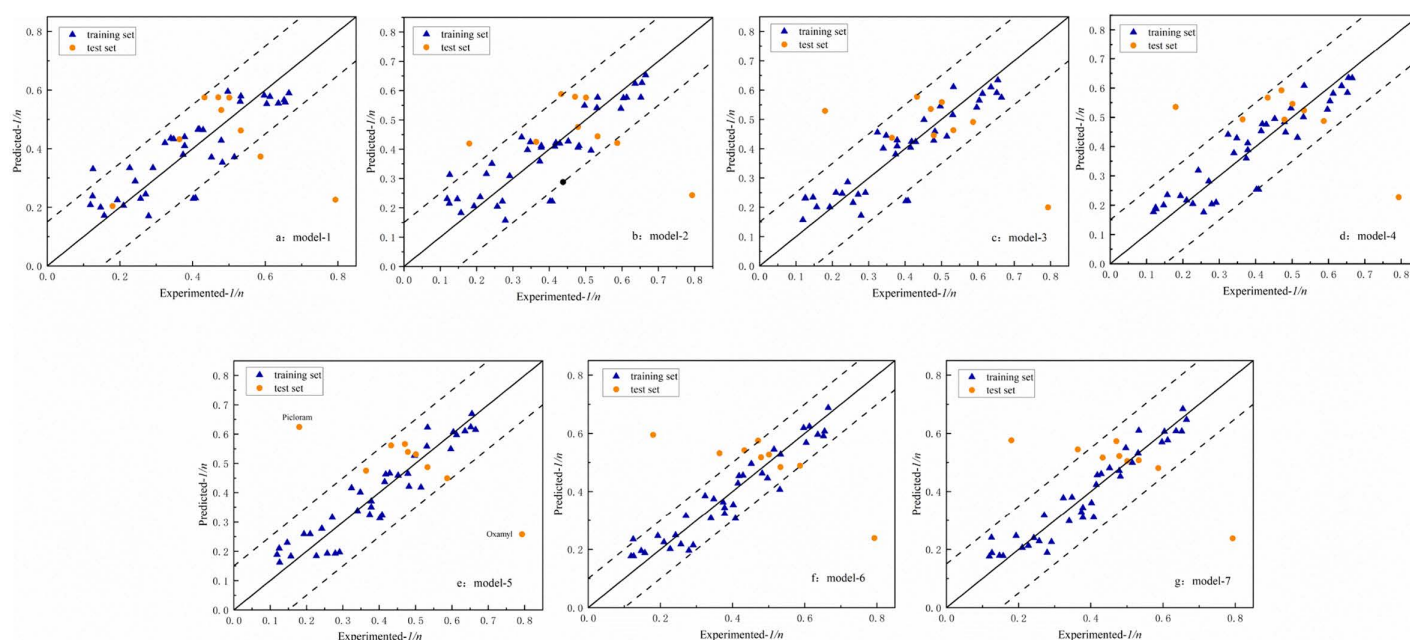


Fig 3. Scatter plot of predicted versus experimental values for models about $1/n$: a) model-1; b) model-2; c) model-3; d) model-4; e) model-5; f) model-6; g) model-7.

<https://doi.org/10.1371/journal.pone.0338483.g003>

present study focused on analyzing parameters such as active site reactivity and electronic effects, thus leading to a large bias in model predictions.

The adsorption models studied by previous authors are listed in Table 4. Most of them focused on the exploration of the hydrophobic properties for a single type of substance, and the parameters chosen were different from those of the present study, which made the compounds with large deviations in the present study.

4 Discussion

4.1 Mechanism explanation

Each quantum chemical parameter in the model plays an important role in the interpretation of the optimal model, and the combined analysis of all descriptors helps to reveal the rules of absorption. The above two models involve different independent variables. For K and $1/n$ in the Freundlich adsorption equation, they cannot completely express the adsorption properties of the adsorbent for the organic compounds independently [41,42]. Hence, two adsorption models based on the Freundlich equation are considered, which is essential to understanding the adsorption mechanism [19,43].

Table 4. Modeling studies in the literature similar to this study.

Research literature (year)	Methods	Type of datasets	Key impact factors
Blum et al. (1994) [8]	MLR	aromatic and aliphatic	molecular connectivity indices
De Ridder et al. (2010) [5]	MLR	organic micropollutants	hydrophobicity (log D), polarizability
Gong et al. (2015) [39]	PLS	mono-nitro aromatic compounds	molecular orbital (E_{LUMO}), the atomic net charges (Q_C^-)
Zhao et al.(2018) [40]	MLR	cationic pharmaceuticals	octanol-water partition coefficient (log P), polar surface area, molecular weight (MW)

Multiple linear regression (MLR), partial-least-squares regression (PLS)

<https://doi.org/10.1371/journal.pone.0338483.t004>

Both the K optimal model and the 1/n optimal model contain Wiberg(C-C)_{min}, and this parameter is used to measure the strength of covalent bonds between carbon and carbon atoms in a molecule. A lower value means that the carbon-carbon bonds in the molecule are weaker, and its structure may be less stable. As can be seen from the regression coefficients in the model, this descriptor is negatively correlated with K and positively correlated with 1/n. From this we can infer that the lower Wiberg(C-C)_{min} of the compound, the more unstable the structure of the molecule is, which may have an effect on the adsorption between the compound molecules and activated carbon. Combined with [S2 Table](#), it can be verified from metolachlor (1.103), cyanazine (1.233), and alachlor (1.104) with the smaller Wiberg (C-C)_{min}. Their corresponding K values (98200, 102000, and 81,700) are higher than the average value (21201). Correspondingly, the corresponding n values (0.125, 0.126, 0.257) are also smaller than the mean value (0.407).

It is noteworthy that in the correlation analysis, K and 1/n both showed a low correlation coefficient with Wiberg(C-C)_{min} (−0.033 and −0.059). Similarly ELUMO (−0.104 and 0.149) and Bond orders(C-C)_{max} (0.306 and −0.291) are also low. In the K-optimal model, the value of the corresponding coefficient for ELUMO is large (−391.288), but the correlation with K in the correlation analysis is small (−0.104). The reason for this aspect may be the complementary effect between the two parameters, where the molecules with lower E_{LUMO} may react with the activated carbon surface through electron transfer, while the weaker Wiberg(C-C)_{min} indicates a decrease in the stability of the bonds in the molecule, which releases the active fragments to enhance adsorption. Thus the synergy between the two explains the co-regulation of the adsorption capacity by electronic effects and structural stability. The synergistic effect between ELUMO and Wiberg(C-C)_{min} refers to their complementary roles in enhancing physisorption. A lower ELUMO value indicates a greater electron-accepting ability, facilitating π–π interactions or weak charge-transfer with the electron-rich graphene layers of activated carbon. Concurrently, a smaller Wiberg(C-C)_{min} reflects weaker intrinsic carbon-carbon bonds within the adsorbate molecule, which may allow for better geometric adaptation and dispersion interactions with the heterogeneous carbon surface. This combined electronic and structural effect optimizes the adsorption affinity, a phenomenon consistent with studies on the adsorption of aromatic and polar compounds onto carbonaceous materials. In the 1/n-optimal model, the corresponding coefficients for Bond orders(C-C)_{max} are negative (−2.775), while those for Wiberg(C-C)_{min} are positive (0.037). Higher Bond orders(C-C)_{max} may inhibit molecular conformational change and limit its ability to fit into the activated carbon pores, and lower Wiberg(C-C)_{min} may promote localized bond breaking or distortion and enhance molecular interactions with adsorption sites. Thus these two parameters regulate the structural dynamics of the molecules during adsorption, thereby affecting the degree of nonlinearity of the isotherms. We may also illustrate this using example compounds. Aldicarb (ELUMO = −0.032 eV, K = 8,270): Its relatively low ELUMO favors electron acceptance from carbon's π-system, enhancing adsorption via charge transfer; Oxamyl (ELUMO = −0.048 eV, K = 1,740): Its higher (less negative) ELUMO indicates poorer electron affinity, reducing charge-transfer potential and adsorption—consistent with its lower K value. In conclusion, analysis of individual variables cannot capture this synergistic effect. Moreover, in subsequent K-model analyses, the Wiberg(C-C)_{min} enhanced robustness (q²: 0.682 → 0.722; SD: 14.305 → 12.171). Within the 1/n model, it refined the nonlinear prediction (R²: 0.885 → 0.917). Its VIF values (K-type: 1.173; 1/n-type: 1.369; [Tables 5–6](#)) confirmed the absence of multicollinearity, justifying its retention.

The electrophilic indices (Fukui(-)_{max}) mainly refers to the active site where the compound is most susceptible to electrophilic reaction, and the active site is the main factor affecting the adsorption effect [44,45]. In the present study, this parameter is inversely proportional to the K value in this model, which means that with higher Fukui(-)_{max} organic compounds have a poor adsorption effect. We can infer that the higher the Fukui(-)_{max} value of the compound, it means that the compound is easy to lose electrons. This may lead to the formation of transient adsorption complexes or quasi-stable adsorption states (e.g., through electrostatic or π–π interactions) during the physisorption process, which could be less favorable for stable adsorption, thereby reducing the overall adsorption performance. For example, the Fukui(-)_{max} of 1,2-dichloroethane (0.3565) was higher than that of 1,2,3-trichloropropane (0.2651), and the K showed that 1,2,3-trichloropropane had a better adsorption capacity.

Table 5. Statistical values of variables for the optimal K-model.

Variable	Regression coefficients	VIF	t	Sig.
$\sum q(O+N)$	-125.215	3.675	-10.982	0.000
$q(CH^+)_{max}$	-695.036	4.138	-7.413	0.000
E_{LUMO}	-391.288	1.531	-5.093	0.000
$Fukui(-)_{max}$	-98.670	1.248	-3.867	0.000
Wiberg(C-C) _{min}	-6.771	1.173	-3.455	0.001
Criteria	—	1-5	>2.179	<0.05

<https://doi.org/10.1371/journal.pone.0338483.t005>

Table 6. Statistical values of variables for the optimal 1/n-model.

Variable	Regression coefficients	VIF	t	Sig.
Bond orders(C-C) _{min}	0.305	2.758	5.821	0.000
Bond orders(C-C) _{max}	-2.775	2.305	-6.382	0.000
$\sum q(-)/N_c$	-3.182	3.512	-5.313	0.000
Wiberg(C-C) _{min}	0.037	1.369	4.072	0.000
$q(C^-)_{max}$	0.717	2.734	3.312	0.002
Criteria	—	1-5	>2.179	<0.05

<https://doi.org/10.1371/journal.pone.0338483.t006>

Bond orders(C-C)_{min} and Bond orders(C-C)_{max} reflect the stability of the molecule. The 1/n-model shows that both are negatively and positively correlated with 1/n, respectively, which reflects that the stability of the chemical bonds influences the degree of nonlinearity of the compounds in the adsorption process. In the adsorption process, smaller Bond orders (C-C) imply weaker carbon-carbon bonds, which makes the compounds easier to be adsorbed by the activated carbon, the closer the adsorption process is to linearity.

According to the correlation (Fig 1), $\sum q(O+N)$ is much more correlated with K than $\sum q(O)$ and $\sum q(N)$, which indicates that the overall charge distribution of the compound molecules has a greater influence on the adsorption effect of activated carbon. For example, diquat has the largest value of $\sum q(O+N)$ (0.1538), while K (2260) is much smaller than the average level; glyphosate has the smallest (-1.5546), while K (87600) is much larger than the average level. $\sum q(-)/N_c$ reflects the overall reflection of the negative charge of the organic compound, and the larger its value, the stronger the overall electronegativity of the compound and the more difficult it is to be adsorbed. This point also explains the relationship that $\sum q(-)/N_c$ is negatively correlated with 1/n. The $q(CH^+)_{max}$ and $q(C^-)_{max}$ reflect the charge distribution of different atoms in organic compounds and molecular polarity [46], which affect the adsorption performance of the adsorbent on the compounds, and hence the linearity of the adsorption isotherm.

Although hydrophobicity parameters (e.g., logP) were not directly introduced in this study, the charge distribution parameters included in the model may also indirectly reflect the association between molecular polarity and hydrophobicity. Combined with S2 Table, it can be found that it may not be possible to accurately assess the adsorption performance of compounds based on the magnitude of a single parameter, while K and n together describe the overall adsorption behavior of the adsorbent on compounds in Freundlich, and therefore the explanation of the adsorption mechanism of activated carbon is valuable only when the K and 1/n are considered together with structure parameters.

In this study, it was found that for compounds containing halogen, nitrogen or oxygen functional groups (e.g., pesticides, pharmaceutical intermediates), the adsorption behavior is more significantly affected by the reactivity of the active site and the electronic effect, although the hydrophobic interaction is crucial for the adsorption mechanism. In the developed model, $Fukui(-)_{max}$ represents the easily attacked active site of the compound, which is capable of directly interacting chemically with the functional groups on the surface of the activated carbon; E_{LUMO} , Bond orders(C-C)_{min} and

Bond orders(C-C)_{min} represent the ability to accept electrons and break chemical bonds, which may affect the charge transfer between adsorbent and compound. The two optimal models developed contain the Wiberg(C-C)_{min}, which reflect the electron density distribution of the bonds and the bond strengths. This finding suggests that the relative importance of reactivity and hydrophobicity parameters may vary with molecular structure differences in complex pollutant systems. The modeling by quantum chemical parameters in this study provides a new perspective for understanding the reactivity-dominated adsorption mechanism.

Take the example of aldicarb and oxamyl. Fukui(-)_{max} corresponds to the sites in its molecule that are easily attacked by oxygen-containing functional groups on the surface of the activated carbon, and higher values allow it to bind to the surface of the activated carbon through hydrogen bonding or electrostatic interaction, enhancing the adsorption capacity. The lower E_{LUMO} indicates that the activated carbon surface may transfer electrons to aldicarb through charge transfer to form an adsorption complex. The magnitude of the parameters of Bond orders(C-C)_{min} and Wiberg(C-C)_{min} is mainly reflected in the C-N bonding in the methylcarbamate group in aldicarb, which is a weak bonding that makes aldicarb susceptible to exchange electrons with the activated carbon surface. For oxamyl, its low Fukui(-)_{max} makes it difficult to form chemical or hydrogen bonds with functional groups on the surface of activated carbon. Higher E_{LUMO} implies low charge transfer efficiency between it and activated carbon, making it difficult to form stable adsorption complexes. Higher Bond orders(C-C)_{min} and Wiberg(C-C)_{min} indicate greater chemical bond strength and more stable electron density distribution in the oxamyl molecules, resulting in a lower possibility of electron exchange or bonding with the activated carbon surface.

4.2 Model validation

To check the stability of the optimal model, t-test and VIF test were conducted for the model, respectively. The statistical values of the independent variables regarding the K optimal model and the 1/n optimal model are shown in [Tables 5](#) and [6](#).

The results show that the values of the sig. of all the respective independent variables in the K optimal model and the 1/n optimal model are less than 0.05, and the absolute value of t is greater than 2.179. For the K optimal model, the order of magnitude of the absolute values of t is $\sum q(O+N) > q(CH^+)_{max} > E_{LUMO} > Fukui(-)_{max} > Wiberg(C-C)_{min}$, and for the 1/n-optimization model, it is $Bond\ orders(C-C)_{min} > Bond\ orders(C-C)_{max} > \sum q(-)/NC > Wiberg(C-C)_{min} > q(C-)_{max}$, which indicates that these variables are acceptable. The VIF is used to test for multicollinearity among the independent variables of a model [\[47\]](#). The criteria for this indicator require that if $VIF > 10$, the model equation is unreliable and needs to be retested or re-modeled; if $VIF = 1-5$, the model equation is acceptable; and if $VIF = 1$, there is no correlation between every variable of the model. [Tables 5](#) and [6](#) show that all the VIF values in the optimal model meet the criteria of 1–5. In addition, the external validation coefficients (K-model: $R_{ext}^2 = 0.699$, $Q_{ext}^2 = 0.668$; 1/n-model: $R_{ext}^2 = 0.738$, $Q_{ext}^2 = 0.738$) of the model all meet the criterion of being greater than 0.5, which suggests that the model has a good predictive potential.

4.3 Y-randomization verification and LMO cross-validation

To check the robustness of the model, Y-randomized validation [\[48\]](#) and Leave-Many-Out cross-validation were performed on the developed model.

For a stable QSAR model, R^2_{yand} and Q^2_{yand} of the Y-randomized model should be lower compared to the original model's R^2 and Q^2 , otherwise, it means that the original model is unreliable. The results are shown in [Fig 4](#), which determined that the two optimal models established in this study have better robustness and stability. The internal prediction ability of the model is also determined based on the comparison of the new cross-validation coefficient (Q_{LMO}^2) with the Q_{LMO}^2 of the original model. The results are shown in [Fig 5](#), the performance of the new LMO model is basically the same as the original model (the optimal model established), so the above two models meet the requirements and passed the validation.

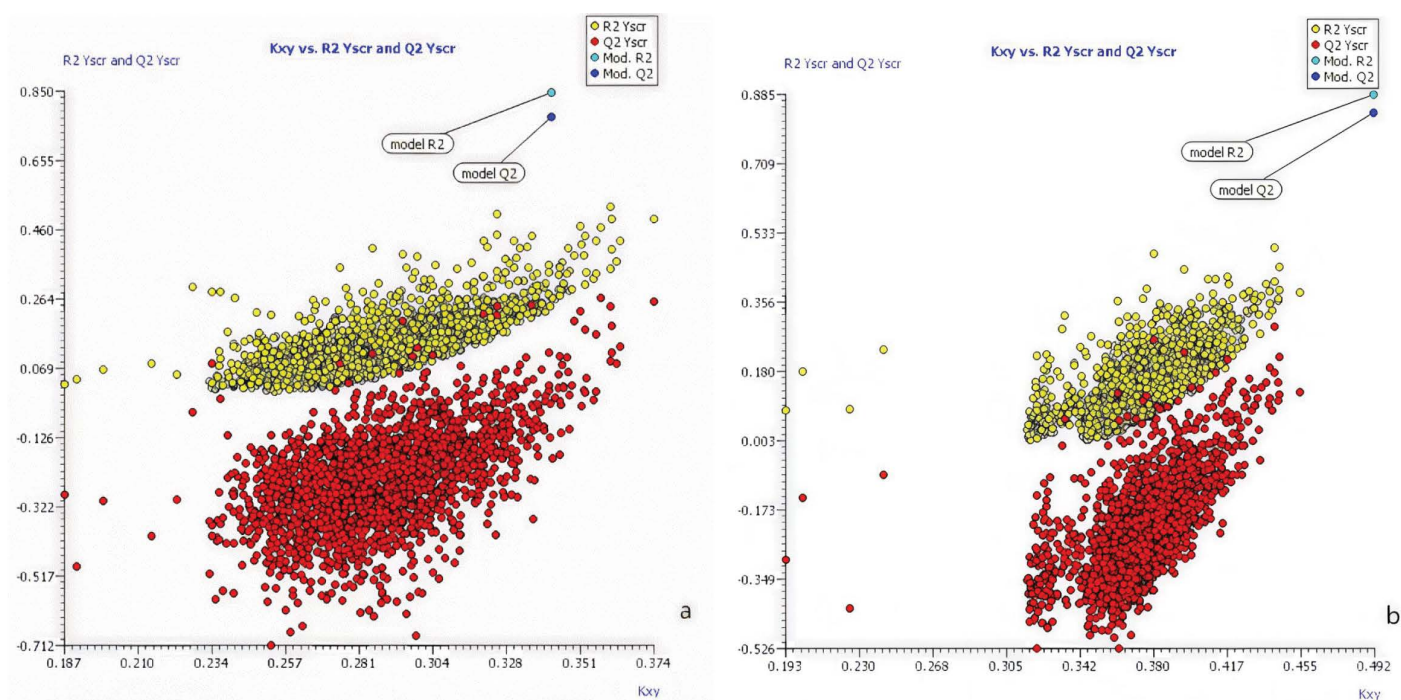


Fig 4. Comparison of Y-randomized model with the optimal model: a) the optimal model for K; b) the optimal model for 1/n.

<https://doi.org/10.1371/journal.pone.0338483.g004>

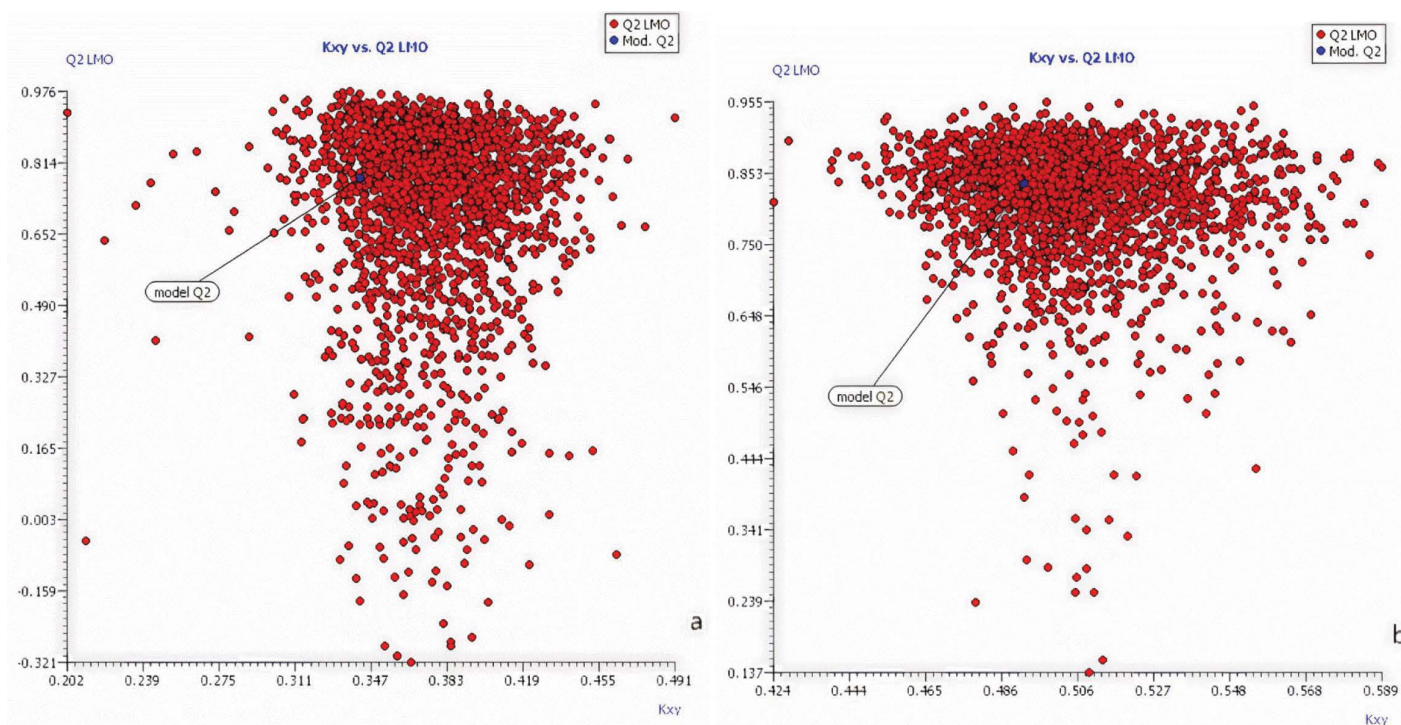


Fig 5. Comparison of the LMO model with the optimal model: a) the optimal model for K; b) the optimal model for 1/n.

<https://doi.org/10.1371/journal.pone.0338483.g005>

4.4 Areas of applicability

To make the models constructed in this study better applied to the study of the adsorption performance of other organic compounds, we also used Williams plots to observe the applicability domains (APDs) of the two optimal models constructed. The details are shown in Fig 6.

The domain of the application test mainly illustrates the range of applicability predicted by the optimal model [49], and only the predictions of the compounds belonging to this field of study are considered to be reliable [50]. The constructed models can predict when the predicted organic compounds satisfy the standard residuals (σ) between $-3, 3$ and the leverage value (h_i) is below the warning bar (h^*) [29,51]. As shown in Fig 6, most of the standard residuals of the organic compounds satisfy the range. However, it was found that the oxamyl in the test set was not in the APD and exceeded the critical value in the K-williams and 1/n-williams plots. The analysis shows that oxamyl is a carbamate compound with a nitrogen-containing heterocyclic structure, which is significantly distinct from the others in structure. This difference may cause its experimental results to be different from the predicted results. In general, the two optimal models still have better stability and prediction ability.

In addition, in order to quantitatively define the model boundaries, we performed Tanimoto similarity analysis based on molecular fingerprints.

As shown in the chemical-space distribution derived from molecular-fingerprint Tanimoto similarities (Fig 7), the MDS plot reveals a relatively well-defined and cohesive chemical space occupied by the majority of the training set compounds, indicating structural consistency within the model's training domain. The test set compounds are generally intermingled with or situated on the periphery of the training set cloud, suggesting that the external validation set shares a reasonable degree of structural similarity with the training compounds, which is crucial for reliable predictive modeling.

4.5 Comparison with models in the literature

According to Table 4, the existing adsorption QSAR models mainly focus on the hydrophobic properties of organic compounds. Only one type of compound is included in the research object, so the developed model may only be applicable to this specific type of compound, and it is difficult to be extended to other types of substances. In addition, the lack of sufficient model performance verification methods may cause problems with the reliability and applicability of the model. This

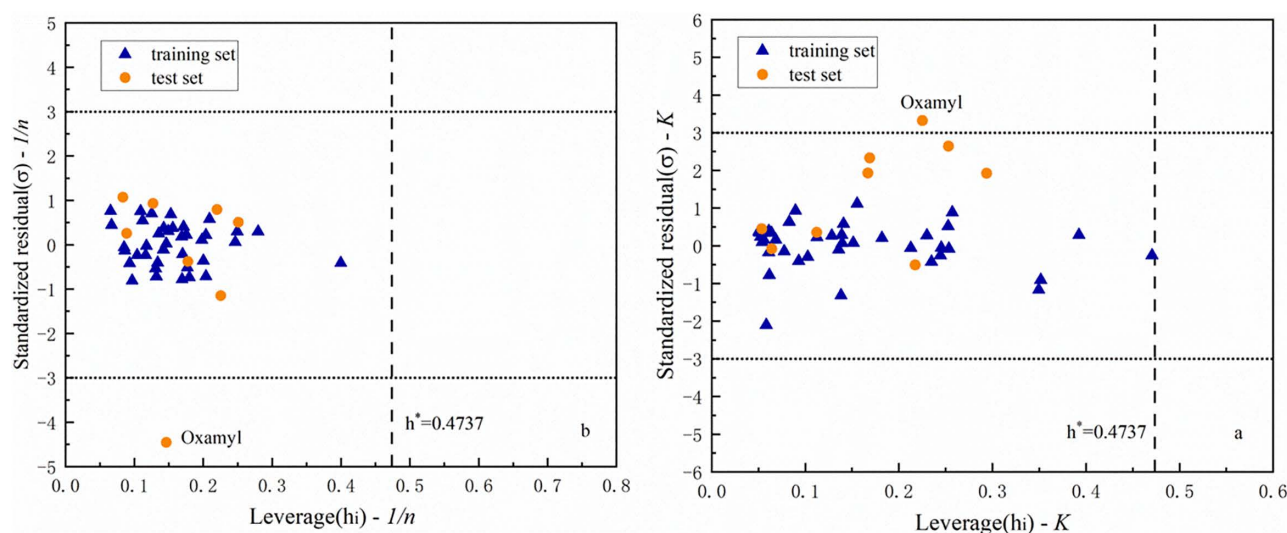


Fig 6. Williams plot of the optimal model: a) the optimal model for K; b) the optimal model for 1/n.

<https://doi.org/10.1371/journal.pone.0338483.g006>

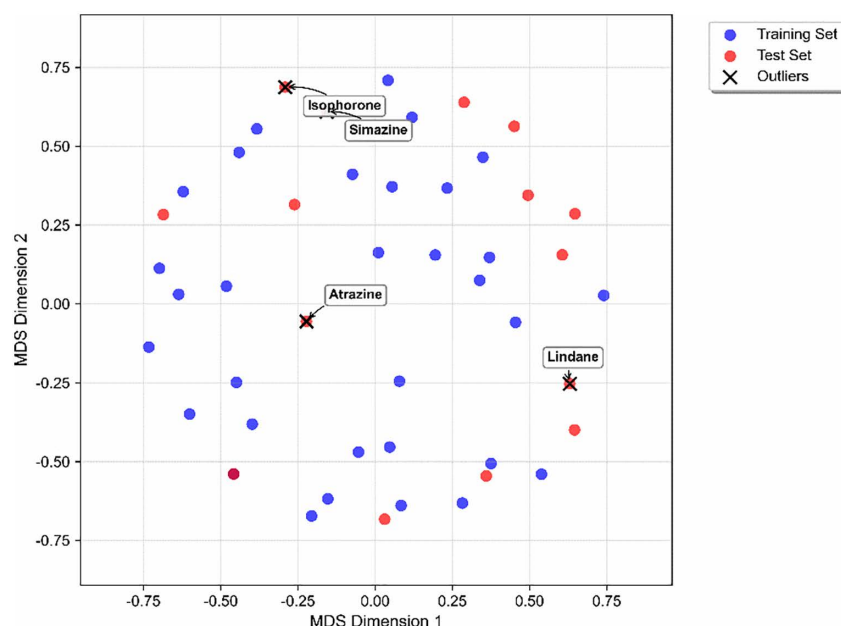


Fig 7. Chemical space distribution based on Tanimoto similarity.

<https://doi.org/10.1371/journal.pone.0338483.g007>

study involved a variety of compounds (including halogenated hydrocarbons, Monoaromatic Hydrocarbons and nitrogen-/oxygen-containing organics), thereby enhancing the applicability of the model.

The second aspect is that time-consuming experiments are normally needed to obtain the exact hydrophobic property of each organic compound, especially for the emerging organics. Actually, the hydrophobic property of each organic compound should also be decided by the molecular structure feature which could normally be quantified through the quantum parameters of the compound. Therefore, another advantage of the QSAR model is the direct connection of organic adsorption characteristics with the molecular structures through the quantum parameters which can be calculated through computer without tedious physical chemistry experiments, especially for the emerging organic compounds.

Furthermore, this model was developed using adsorption data exclusively from Fitrasorb-400 activated carbon, which serves as a well-characterized reference material. The identified QSARs capture the fundamental influence of adsorbate properties (electronic and structural) on adsorption affinity. While the absolute adsorption capacity (K value) can be influenced by the surface heterogeneity of different activated carbons (e.g., oxygen-containing functional groups, ash content), the underlying trends described by the quantum chemical descriptors are expected to be qualitatively transferable to carbons with similar non-specific surface domains. To quantitatively predict adsorption onto carbons with vastly different surface chemistries, future extensions of this framework will incorporate material-specific descriptors (e.g., surface O/C ratio, point of zero charge) into a comprehensive model.

4.6 Limitations and future perspectives

It is important to note the limitations of the current model. The experimental data were obtained under natural aqueous conditions (pH 5.3–8.0) without adjustment, and the descriptors were calculated for neutral molecular species. Consequently, the model's predictive ability for strongly ionizable emerging contaminants (e.g., metalaxyl and simazine) under different pH conditions may be limited. Future work will focus on integrating pH-corrected molecular descriptors to significantly expand the applicability domain of the QSAR model.

5 Conclusion

The modeling results showed that the two key parameters (K and $1/n$) of the Freundlich Isotherm are correlated well with certain quantum chemical parameters which indicates that the applicability of the empirical Isotherm is actually related to the molecular structure characteristics of the organic compounds. The specific quantum parameters determining the value of K were discovered as $\sum q(O+N)$, $q(CH^+)_{max}$, E_{LUMO} , $Fukui(-)_{max}$, $Wiberg(C-C)_{min}$, which indicates that the charge distribution, carbon bond energy and active site energy in the molecular structure are the main factors affecting the organic adsorption efficiency by activated carbon. The QSAR model of $1/n$ exhibited the similar finding that $1/n$ value is also correlated with the molecular structure characteristics.

The two optimal models were further confirmed to be stable, robust, and accurate by the standard evaluation. The two QSAR models can be used to predict the adsorption efficiency of other similar compounds (subject to model applicability domain testing) based on the related quantum parameters. Therefore, this study provides an important and convenient reference method for evaluating the applicability of activated carbon adsorption unit to control various emerging organic pollutants in drinking water plants and also provides them a basis for constructing an intelligence management system.

Future work will extend the QSAR framework to multi-solute systems by incorporating competitive adsorption coefficients and molecular dynamics simulations.

Supporting information

S1 Table. Meaning of parametric descriptors of compounds.

(DOCX)

S2 Table. Information on Freundlich parameters and 31 molecular parameters for 47 organic compounds.

(DOCX)

S3 Table. Meaning of evaluation indexes.

(DOCX)

Acknowledgments

This work gratefully acknowledges Prof. Paola Gramatica for providing a free license of the QSARINs software (website: www.qsar.it). We also extend our sincere thanks to Shanghai Jiao Tong University for their generous sharing of expertise and experience in the use of QSAR software. Finally, we extend our gratitude to all the reviewers of this paper for their constructive suggestions for improvement.

Author contributions

Conceptualization: Ting Tan, Qunshan Wei, Zhemin Shen, Qiong Liu, Xinshan Song, Yuhui Wang, Charles Nzila, Christopher W.K. Chow.

Data curation: Ting Tan, Qunshan Wei.

Formal analysis: Ting Tan.

Funding acquisition: Zhemin Shen, Yuhui Wang.

Investigation: Ting Tan, Zhemin Shen, Qiong Liu, Xinshan Song, Charles Nzila, Christopher W.K. Chow.

Methodology: Qunshan Wei, Zhemin Shen, Qiong Liu, Xinshan Song.

Project administration: Zhemin Shen.

Resources: Qunshan Wei, Zhemin Shen.

Software: Zhemin Shen.

Supervision: Zhemin Shen, Qiong Liu, Xinshan Song, Charles Nzila, Christopher W.K. Chow.

Validation: Ting Tan.

Visualization: Ting Tan.

Writing – original draft: Ting Tan.

Writing – review & editing: Siyi Ding, Yefan Yang, Qunshan Wei, Charles Nzila, Christopher W.K. Chow.

References

1. Daud MK, Nafees M, Ali S, Rizwan M, Bajwa RA, Shakoor MB, et al. Drinking Water Quality Status and Contamination in Pakistan. *Biomed Res Int*. 2017;2017:7908183. <https://doi.org/10.1155/2017/7908183> PMID: [28884130](#)
2. Lyu J, Chen Y, Zhang L. Antibiotics in Drinking Water and Health Risks - China, 2017. *China CDC Wkly*. 2020;2(23):413–7. <https://doi.org/10.46234/ccdcw2020.106> PMID: [34594669](#)
3. Polishchuk P. Interpretation of Quantitative Structure-Activity Relationship Models: Past, Present, and Future. *J Chem Inf Model*. 2017;57(11):2618–39. <https://doi.org/10.1021/acs.jcim.7b00274> PMID: [28949520](#)
4. Sharma P, Ranjan P, Chakraborty T. Applications of conceptual density functional theory in reference to quantitative structure–activity / property relationship. *Molecular Physics*. 2024;122(23). <https://doi.org/10.1080/00268976.2024.2331620>
5. de Ridder DJ, Villacorte L, Verliefde ARD, Verberk JQJC, Heijman SGJ, Amy GL, et al. Modeling equilibrium adsorption of organic micropollutants onto activated carbon. *Water Res*. 2010;44(10):3077–86. <https://doi.org/10.1016/j.watres.2010.02.034> PMID: [20236679](#)
6. Yang K, Zhu L, Xing B. Adsorption of polycyclic aromatic hydrocarbons by carbon nanomaterials. *Environ Sci Technol*. 2006;40(6):1855–61. <https://doi.org/10.1021/es052208w> PMID: [16570608](#)
7. Hobbs JA, Lindner A, Mazyck D. Adsorption of substituted aromatic compounds by powdered activated carbon: A mechanistic approach to quantitative structure–activity relationships. 2005.
8. Blum DJW, Suffet IH, Duguet JP. Quantitative structure–activity relationship using molecular connectivity for the activated carbon adsorption of organic chemicals in water. *Water Research*. 1994;28(3):687–99. [https://doi.org/10.1016/0043-1354\(94\)90149-x](https://doi.org/10.1016/0043-1354(94)90149-x)
9. Rao BH, Asolekar SR. QSAR models to predict effect of ionic strength on sorption of chlorinated benzenes and phenols at sediment–water interface. *Water Res*. 2001;35(14):3391–401. [https://doi.org/10.1016/S0043-1354\(01\)00053-7](https://doi.org/10.1016/S0043-1354(01)00053-7) PMID: [11547860](#)
10. Dewulf J, Dewettinck T, De Visscher A, Van Langenhove H. Sorption of chlorinated C1- and C2-hydrocarbons and monocyclic aromatic hydrocarbons on sea sediment. *Water Research*. 1996;30(12):3130–8. [https://doi.org/10.1016/S0043-1354\(96\)00205-9](https://doi.org/10.1016/S0043-1354(96)00205-9)
11. Mader BT, Uwe-Goss K, Eisenreich SJ. Sorption of Nonionic, Hydrophobic Organic Chemicals to Mineral Surfaces. *Environ Sci Technol*. 1997;31(4):1079–86. <https://doi.org/10.1021/es960606g>
12. Kamlet MJ, Doherty RM, Abraham MH, Marcus Y, Taft RW. Linear solvation energy relationship. 46. An improved equation for correlation and prediction of octanol/water partition coefficients of organic nonelectrolytes (including strong hydrogen bond donor solutes). *J Phys Chem*. 1988;92(18):5244–55. <https://doi.org/10.1021/j100329a035>
13. Patel HM, Noolvi MN, Sharma P, Jaiswal V, Bansal S, Lohan S, et al. Quantitative structure–activity relationship (QSAR) studies as strategic approach in drug discovery. *Med Chem Res*. 2014;23(12):4991–5007. <https://doi.org/10.1007/s00044-014-1072-3>
14. Miltner RJ, Love OT. Comparison of procedures to determine adsorption capacity of volatile organic compounds on activated carbon. 1985.
15. Speth TF, Miltner RJ. Technical note: adsorption capacity of GAC for synthetic organics. *Journal AWWA*. 1998;90(4):171–4. <https://doi.org/10.1002/j.1551-8833.1998.tb08420.x>
16. Randtke SJ, Snoeyink VL. Evaluating GAC adsorptive capacity. *Journal AWWA*. 1983;75(8):406–13. <https://doi.org/10.1002/j.1551-8833.1983.tb05177.x>
17. Hamadache M, Benkortbi O, Hanini S, Amrane A, Khaouane L, Si Moussa C. A Quantitative Structure Activity Relationship for acute oral toxicity of pesticides on rats: Validation, domain of application and prediction. *J Hazard Mater*. 2016;303:28–40. <https://doi.org/10.1016/j.jhazmat.2015.09.021> PMID: [26513561](#)
18. Lu T, Chen F. Multiwfn: a multifunctional wavefunction analyzer. *J Comput Chem*. 2012;33(5):580–92. <https://doi.org/10.1002/jcc.22885> PMID: [22162017](#)
19. El-Desouky MG, Khalil MAG, El-Affify MAM, El-Bindary AA, El-Bindary MA. Effective methods for removing different types of dyes – modelling analysis, statistical physics treatment and DFT calculations: a review. *Desalination and Water Treatment*. 2022;280:89–127. <https://doi.org/10.5004/dwt.2022.29029>
20. Liu Y, Cheng Z, Liu S, Tan Y, Yuan T, Yu X, et al. Quantitative structure activity relationship (QSAR) modelling of the degradability rate constant of volatile organic compounds (VOCs) by OH radicals in atmosphere. *Sci Total Environ*. 2020;729:138871. <https://doi.org/10.1016/j.scitoenv.2020.138871> PMID: [32361444](#)

21. Jiang A, Cheng Z, Shen Z, Guo W. QSAR study on the removal efficiency of organic pollutants in supercritical water based on degradation temperature. *Chem Cent J*. 2018;12(1):16. <https://doi.org/10.1186/s13065-018-0380-y> PMID: 29442196
22. Gramatica P, Chirico N, Papa E, Cassani S, Kovarich S. QSARINS: A new software for the development, analysis, and validation of QSAR MLR models. *J Comput Chem*. 2013;34(24):2121–32. <https://doi.org/10.1002/jcc.23361>
23. Gramatica P, Cassani S, Chirico N. QSARINS-chem: Insubria datasets and new QSAR/QSPR models for environmental pollutants in QSARINS. *J Comput Chem*. 2014;35(13):1036–44. <https://doi.org/10.1002/jcc.23576> PMID: 24599647
24. Ouaisa M, Laidi M, Benkortbi O, Maarof H. QSPR modeling to predict surface tension of psychoanaleptic drugs using the hybrid DA-SVR algorithm. *J Mol Graph Model*. 2025;134:108896. <https://doi.org/10.1016/j.jmgm.2024.108896> PMID: 39476630
25. Cheng Z, Yang B, Chen Q, Gao X, Tan Y, Yuan T, et al. Quantitative-Structure-Activity-Relationship (QSAR) models for the reaction rate and temperature of nitrogenous organic compounds in supercritical water oxidation (SCWO). *Chemical Engineering Journal*. 2018;354:12–20. <https://doi.org/10.1016/j.cej.2018.07.167>
26. Bhuvaneswari S, Aakash VB, Ramalakshmi N, Arunkumar S. Quantitative Structure–Activity Relationship Analysis and Validation of New DNA Gyrase Inhibitors. *Pharm Chem J*. 2021;55(9):886–907. <https://doi.org/10.1007/s11094-021-02513-x>
27. Rücker C, Rücker G, Meringer M. y-Randomization and its variants in QSPR/QSAR. *J Chem Inf Model*. 2007;47(6):2345–57. <https://doi.org/10.1021/ci700157b> PMID: 17880194
28. Clark RD. Boosted leave-many-out cross-validation: the effect of training and test set diversity on PLS statistics. *J Comput Aided Mol Des*. 2003;17(2–4):265–75. <https://doi.org/10.1023/a:1025366721142> PMID: 13677492
29. Király P, Kiss R, Kovács D, Ballaj A, Tóth G. The Relevance of Goodness-of-fit, Robustness and Prediction Validation Categories of OECD-QSAR Principles with Respect to Sample Size and Model Type. *Mol Inform*. 2022;41(11):e2200072. <https://doi.org/10.1002/minf.202200072> PMID: 35773201
30. Gramatica P. Principles of QSAR models validation: internal and external. *QSAR Comb Sci*. 2007;26(5):694–701. <https://doi.org/10.1002/qsar.200610151>
31. Wiberg KB. Application of the pople-santry-segal CNDO method to the cyclopropylcarbiny and cyclobutyl cation and to bicyclobutane. *Tetrahedron*. 1968;24(3):1083–96. [https://doi.org/10.1016/0040-4020\(68\)88057-3](https://doi.org/10.1016/0040-4020(68)88057-3)
32. Osman Abdelkarim OI, Asiri AM. NBO technique as a descriptor of aromaticity. *Computational and Theoretical Chemistry*. 2022;1210:113637. <https://doi.org/10.1016/j.comptc.2022.113637>
33. Li H, Wang X, Li Z, Pan S. Theoretical Study on the Structure and Properties of Graphdiyne. *Acta Chim Sinica*. 2013;71(1):75. <https://doi.org/10.6023/a12090728>
34. Mabkhot YN, Aldawsari FD, Al-Showiman SS, Barakat A, Soliman SM, Choudhary MI, et al. Novel enaminone derived from thieno [2,3-b] thiene: Synthesis, x-ray crystal structure, HOMO, LUMO, NBO analyses and biological activity. *Chem Cent J*. 2015;9:24. <https://doi.org/10.1186/s13065-015-0100-9> PMID: 26064187
35. Razali NZK, Wan Hassan WNS, Sheikh Mohd Ghazali SAI, Mohd Shotor SN, Dzulkifli NN. DFT, Fukui indices, and molecular dynamic simulation studies on corrosion inhibition characteristics: a review. *Chem Pap*. 2023;78(2):715–31. <https://doi.org/10.1007/s11696-023-03148-w>
36. Zhu H, Guo W, Shen Z, Tang Q, Ji W, Jia L. QSAR models for degradation of organic pollutants in ozonation process under acidic condition. *Chemosphere*. 2015;119:65–71. <https://doi.org/10.1016/j.chemosphere.2014.05.068> PMID: 24972172
37. Tropsha A, Golbraikh A. Predictive QSAR modeling workflow, model applicability domains, and virtual screening. *Curr Pharm Des*. 2007;13(34):3494–504. <https://doi.org/10.2174/138161207782794257> PMID: 18220786
38. Tropsha A. Best Practices for QSAR Model Development, Validation, and Exploitation. *Mol Inform*. 2010;29(6–7):476–88. <https://doi.org/10.1002/minf.201000061> PMID: 27463326
39. Wang R, Jiang J, Pan Y, Cao H, Cui Y. Prediction of impact sensitivity of nitro energetic compounds by neural network based on electrotopological-state indices. *J Hazard Mater*. 2009;166(1):155–86. <https://doi.org/10.1016/j.jhazmat.2008.11.005> PMID: 19101083
40. Chiou CT. Partition and adsorption of organic contaminants in environmental systems. John Wiley & Sons. 2003.
41. Weber WJ, Digiano FA. Process dynamics in environmental systems. 1996.
42. Khan K, Roy K, Benfenati E. Ecotoxicological QSAR modeling of endocrine disruptor chemicals. *J Hazard Mater*. 2019;369:707–18. <https://doi.org/10.1016/j.jhazmat.2019.02.019> PMID: 30831523
43. Liu Y, Liu S, Cheng Z, Tan Y, Gao X, Shen Z, et al. Predicting the rate constants of volatile organic compounds (VOCs) with ozone reaction at different temperatures. *Environ Pollut*. 2021;273:116502. <https://doi.org/10.1016/j.envpol.2021.116502> PMID: 33486251
44. Pal S, Chandrakumar KRS. Critical Study of Local Reactivity Descriptors for Weak Interactions: Qualitative and Quantitative Analysis of Adsorption of Molecules in the Zeolite Lattice. *J Am Chem Soc*. 2000;122(17):4145–53. <https://doi.org/10.1021/ja993617t>
45. Ren Y, Tan Y, Cheng Z, Liu Y, Liu S, Shen Z, et al. QSAR model and mechanism research on color removal efficiency of dyeing wastewater by FeCl₃ coagulation. *Ecotoxicol Environ Saf*. 2022;240:113693. <https://doi.org/10.1016/j.ecoenv.2022.113693> PMID: 35653976
46. Gupta S, Basant N. Modeling the reactivity of ozone and sulphate radicals towards organic chemicals in water using machine learning approaches. *RSC Adv*. 2016;6(110):108448–57. <https://doi.org/10.1039/c6ra22865h>

47. Kiralj R, Ferreira MMC. Basic validation procedures for regression models in QSAR and QSPR studies: theory and application. *J Braz Chem Soc.* 2009;20(4):770–87. <https://doi.org/10.1590/s0103-50532009000400021>
48. Gadaleta D, Mangiatordi GF, Catto M, Carotti A, Nicolotti O. Applicability Domain for QSAR Models. *International Journal of Quantitative Structure-Property Relationships.* 2016;1(1):45–63. <https://doi.org/10.4018/ijqspr.2016010102>
49. Melagraki G, Afantitis A. Enalos InSilicoNano platform: an online decision support tool for the design and virtual screening of nanoparticles. *RSC Adv.* 2014;4(92):50713–25. <https://doi.org/10.1039/c4ra07756c>
50. Hodges JS, Atkinson AC. Plots, Transformations, and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis. *Journal of the American Statistical Association.* 1987;82(398):689. <https://doi.org/10.2307/2289498>
51. Speth TF, Miltner RJ. Technical Note: Adsorption Capacity of GAC for Synthetic Organics. *Journal AWWA.* 1990;82(2):72–5. <https://doi.org/10.1002/j.1551-8833.1990.tb06922.x>