Desalination and Water Treatment
www.deswater.com

doi:10.5004/dwt.2017.21039

87 (2017) 257–267
August

QSAR models for the degradation of organic compounds in three pH conditions of potassium permanganate system

Zhiwen Cheng, Zhemin Shen[a,*], Wenchao Ji, Qingli Tang, Bowen Yang, Pingru Su

School of Environmental Science and Engineering, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai 200240, China, email: chengzhiwen@sjtu.edu.cn (Z. Cheng), zmshen@sjtu.edu.cn (Z. Shen), wenchaoji@sjtu.edu.cn (W. Ji), tangqingli@sjtu.edu.cn (Q. Tang), yangbowen1988@sjtu.edu.cn (B. Yang), supingru@sjtu.edu.cn (P. Su)

Received 12 March 2017; Accepted 13 June 2017

ABSTRACT

Potassium permanganate oxidation is one of the effective water treatment processes applied to remove organic compounds, but structure variances of organic molecules and different water conditions result in different effects. To provide a complete understanding of potassium permanganate oxidation, it is meaningful to develop some relationships between reaction rate constants of organic compounds and structure of molecules in three conditions (acid, neutral and alkaline). In this study, 22 diverse organic compounds were measured for the reaction rate and Quantitative Structure Activity Relationship (QSAR) models were developed based on reaction rate constants of organic substances and 17 quantum descriptors. Quantum chemical descriptors were obtained by using Gaussian 09 and Material Studio 6.1, including μ, E_{B3LYP}, $q(CH^+)$, $q(C^-)$, $q(H^+)$, E_{LUMO}, E_{HOMO}, bond order and Fukui indices. A set of 18 compounds were used as training set to develop models and 4 compounds were used as test set for external validation. Three optimal models were selected in three conditions, respectively. In the models, energy of the highest occupied molecular orbital (E_{HOMO}) and Fukui indices appeared in all the conditions, indicating the two quantum descriptors play the important role during the oxidation process. Based on the evaluation criteria, model 4 in the acid condition was recommended and exhibited optimum stability and predictive ability. The equation of model 4 is $\ln k' = 0.53 + 0.22\ \mu + 14.26\ E_{HOMO} + 4.89\ q(H^+) - 18.21\ f(0)_x$, the results also indicate that the dipole moment (μ) is much more closely related to the value of $\ln k'$ than other quantum descriptors in potassium permanganate oxidation process.

Keywards: Organic compounds; Potassium permanganate oxidation; QSAR; Quantum chemical descriptors; Reaction rate constants; pH condition

1. Introduction

A variety of organic pollutants have been increasingly emerged in wastewater, the composition of organic pollutions is complicated and diverse, the pollutants also increase difficulties to degrade them by biological, physical, and chemical treatment technology [1–3] because of their complicated structures, including double bonds, activated aromatic rings, specific ring atoms, etc. [4,5]. Increasing attention has been paid to the advanced oxidation processes (AOPs) wastewater treatment methods in recent years, due to its simple operation, fast reaction kinetics and high efficiency for pollution control [6]. But the AOPs mainly depend on the hydroxyl radical which is produced by the oxidant during the reaction process, and it is stricter for the reaction condition. Compared with AOPs in water treatment, potassium permanganate, as a metal–oxo reagent, has less limitation as an oxidant, and may be more practical as an addible oxidant and does not apparently rely on generating a hydroxyl radical to oxidize the organic compounds, and has higher efficiency in water treatment. Former researchers focused on more than a century to indicate that the double carbon-carbon bond could be attacked powerfully by metal–

*Corresponding author.

oxo reagents [7]. Schnarr et al. [8] observed that more than 90% of the TCE and PCE were degraded after using aqueous permanganate. Pang et al. [9] found that the most widely used brominated flame retardants tetrabromobisphenol A (TBrBPA) could be readily transformed to several dimeric products by potassium permanganate. Rachel and Paul [10] explored the kinetics of particular contaminants by permanganate, their results suggested that the MnO_4^- had great influence on the 24 contaminants they selected. In addition, it is well known that the structure of a molecule greatly influences the degradation of a compound. Ruppert et al. [11] found out that the structures of some organic pollutants had obviously influenced the results during the way they were mineralized by OH-radicals. Ren et al. [12] studied the degradation reactivity mechanism of 116 diverse compounds by using structure descriptors. It is of great significance to study the degradation of compounds by using their structures.

Quantitative Structure Activity Relationship (QSAR) analysis as a theoretical predicted method has the advantages of rapid and cost-effective, it is used to alternate to traditional analytical methods and has attracted great attentions in the last several years [13,14]. It provides intuitively understanding of the relationship between degradation behaviors and structure of a molecule. Moreover, quantum chemical descriptors influence the effects of degradation indeed and have been considered recently [15]. Energy of the highest occupied molecular orbital (E_{HOMO}) and energy of the lowest unoccupied molecular orbital (E_{LUMO}) were reported to influence the oxidation at the molecular orbital level [16]. Fukui indices were defined as the derivative of electron density with respect to the number of electrons at constant molecular geometry, and results showed that Fukui indices were the crucial region selectivity indicators in chemical reactions [17,18]. It is meaningful to set up some relationships between the degradation behaviors and structure of molecule by using quantum chemical descriptors. In fact, several QSAR models were reported to estimate this relationship, Li et al. [19] established a model by using 14 quantum chemical descriptors, dragon descriptors and structural fragments. Zhu et al. [20] developed some models by using 17 quantum chemical descriptors in ozonation process. Jia et al. [21] built a model by 13 molecular descriptors in Fenton process. But QSAR models for the degradation of organic compounds have seldom been developed in potassium permanganate oxidation, and the relationship between kinetics and the structure of molecules in potassium permanganate system is still rarely reported.

In this work, experiments were carried out for the kinetics on 22 different kinds of organic compounds with various structures in three conditions and QSAR models were developed for the relationship between the reaction rate constants and quantum chemical descriptors. It is our emphasis to develop appropriate models in different conditions, estimate the reaction rates of organic compounds in potassium permanganate system, explore the difference of degradation mechanism among three conditions and find out the best condition for the degradation in potassium permanganate system.

2. Experimental and computation methods

2.1. Experimental methods

22 widely used organic compounds were selected as experimental materials, including rhodamine B, methylene blue, nitrobenzene and other different compounds. They were used to study the relationship between reaction rate constants and the structural descriptors. All experiments were conducted in 2000 mL reactors at the temperature of 298.15 K at pH = 3 ± 0.1, pH = 7 ± 0.1 and pH = 10 ± 0.1. The initial concentration of each organic compound solution was 100 mg/L. Potassium permanganate (AR, Sinopharm Chemical Reagent Co., Ltd.) was used as oxidant. The mole ratio of each organic compound and potassium permanganate was 1:20, and saturated solution of sodium sulfite was used for terminating the reaction. All other reagents were analytical pure. During the reaction process, we detected the concentrations of each organic compound at different residence time. Firstly, the samples were filtered through a membrane of 0.45 μm filters to remove the manganese dioxide precipitates which was formed in the reaction process. Then UV spectrophotometer (UV-1600, Shanghai Mapada Instruments Co., Ltd.) was used to analyze the concentrations changes of organic compounds at their maximum absorption wavelengths. At last, the reaction rate constants were obtained from the chemical reaction rate equation.

2.2. Computation details

All calculations of the 22 organic compounds were carried out by Gaussian 09 (DFT B3LYP/6-311G level) and Material Studio 6.1 (Dmol3/GGA-BLYP/DNP (3.5) basis). The total energy (E_{B3LYP}) calculations of the optimized geometries were based on B3LYP method. The structures of 22 chemicals were optimized based on density functional theory (DFT) B3LYP/6-311G method using Gaussian 09. The frequency calculations were performed to ensure they were at the minimum potential energy surface. Exchange and correlation terms were considered with a B3LYP function (6-311G basis set). Then the natural population analysis (NPA) of atomic charge was obtained by the same method, at last, the values of quantum descriptors were obtained from the Gaussian 09 output files. The quantum chemical descriptors include dipole moment (μ), the total energy of a molecule E_{B3LYP}, most positive partial charge on a hydrogen atom (qH^+), minimum and maximum negative partial charge on a carbon atom ($q(C^-)_n/q(C^-)_x$), minimum and maximum positive partial charge on a hydrogen atom linked with a carbon atom ($q(CH^+)_n/q(CH^+)_x$), E_{HOMO} and E_{LUMO}.

Other descriptors including bond order (BO) and Fukui indices (f(+), f(–) and f(0)) were calculated by Material Studio 6.1. Bond order is the number of chemical bonds between a pair of atoms, suggesting the stability of a bond. Molecule tends to be more stable if the BO is larger when BO is smaller than 4 [22]. BO_x and BO_n refer to the maximum and minimum of bond order values in the molecule, respectively. Fukui indices are defined as the derivative of the electron density with respect to the number of electrons at constant molecular geometry [23,24]. They are key factors to describe the decomposition sequence of molecular structure in the oxidation process.

Eqs. (1)–(3) are three main basis points of Fukui indices.

$$f(+) = \rho N+1(r) - \rho N(r) \quad (1)$$

$$f(-) = \rho N(r) - \rho N-1(r) \quad (2)$$

$$f(0) = \frac{1}{2}\left[\rho N+1(r) - \rho N-1(r)\right] \quad (3)$$

where ρ*N+1(r)*, ρ*N(r)*, ρ*N-1(r)* are the electron densities of the *N+1, N, N-1* electron system, respectively.

Fukui indices ($f(+)_x$, $f(-)_x$ and $f(0)_x$) express the affinity with nucleophilic attack, electrophilic attack and radical attack, $f(+)_n$, $f(-)_n$ and $f(0)_n$ stand for the minimum values on main-chain carbon atom, they are significant for analysis of site reactive selectivity among the oxidation paths. All calculations and localized double numerical basis set with polarization functional (DNP) were adopted, as implemented in DMol3 code in the Material Studio 6.1. The self-consistent field procedure was carried out with a convergence criterion of 10^{-6} Ha. on energy and electron density. Density mixing was set at 0.2 charge and 0.5 spin. The smearing of electronic occupations was set as 5×10^{-3} Ha. 22 kinds of compounds were geometry optimized with the same setup.

2.3. QSAR model construction and validation methods

Stepwise regression procedure was used to build QSAR models by SPSS 20.0 for Windows software program. The quality of developed QSAR models were evaluated in accordance with the squared regression coefficient (R^2), the standard deviation (SD), the t test, the Fisher test and Root Mean Square Error (RMSE). To validate the models, both of the internal validation (q^2) and the external validation (Q^2_{EXT}) were performed by the leave one out (LOO) cross-validation. Y- Randomization validation and Υ^2_m validation were also carried out to validate the models.

3. Results and discussion

3.1. Experiment results

The potassium permanganate degradation process of 22 organic compounds can be illustrated as equation below.

Organic compounds + Potassium permanganate → Ps (products and intermediates) (4)

Based on Eq. (4), the potassium permanganate degradation rate equation can be explained as follows

$$\frac{-dC_t}{dt} = kC_t^{\ m}C_p^{\ n} \quad (5)$$

where C_t (mol/L) and C_p (mol/L) are the concentrations of organic compounds and potassium permanganate in aqueous solution, respectively; *t* is the reaction time; *k* is the reaction rate constant, *m* and *n* are the reaction orders of organic compounds and potassium permanganate, respectively. Since the concentration of potassium permanganate is sufficient during the oxidation process, it can be regarded as a constant, assuming that C_p has no influence on the potassium permanganate diffusion rate under stirring in aqueous solution. Thus Eq. (5) can be simplified as Eq. (6).

$$\frac{-dC_t}{dt} = k'C_t^{\ m} \quad (6)$$

where *k'* is an apparent reaction rate constants and *m* is the total reaction order.

When the total reaction order *m* is one, the reaction equation can be shown as Eq. (7)

$$ln\left(\frac{C_0}{C_t}\right) = k't \quad (7)$$

where C_0 is the initial concentration of organic compounds in the reaction system

The relationship between the reaction rate constants of colour removal and residence time of 22 organic compounds were investigated during the oxidation process, the *k'* value of each compound was calculated from the linear regression of $\ln(C_0/C_t)$ versus reaction time. In all conditions, the plots of $\ln(C_0/C_t)$ versus time were linear, indicating that the reaction rate followed the first-order reaction. The results are shown in the following figures and *k'* values are listed in Table 1. Fig. 1a shows the reaction rate constants in acid condition, from the two figures, we can easily find that *k'* value of each compound is distinguished

Table 1
k' values of 22 organic compounds in three pH conditions

Compounds	*k'*		
	pH = 3	pH = 7	pH = 10
Phloroglucinol	0.094	0.026	0.073
Dibutyl phthalate	0.036	0.032	0.167
Rhodamine B	2.712	0.098	0.456
Isatin	0.098	0.082	0.102
Methylene blue	1.267	0.342	0.385
Eriochrome blue black R	0.482	1.912	0.577
Metanil yellow	0.892	1.985	1.300
Dimethyl phthalate	0.042	0.066	0.054
Bromophenol blue	0.970	0.866	0.525
o-phthalic acid	0.256	0.123	0.179
Crystal violet	2.355	0.797	2.178
3,4-dichloroaniline	2.168	0.734	0.412
Basic fuchsin	3.327	0.699	2.972
Nitrobenzene	0.135	0.126	0.040
Orange G	0.367	2.501	0.572
Methyl orange	0.683	1.150	1.354
2,4-dichlorophenol	0.091	0.860	0.095
Azure i	2.199	0.849	0.277
Diethyl phthalate	0.026	0.124	0.224
5-chloro-2-methylaniline	0.071	0.093	0.124
Indigo	1.355	0.543	2.968
Aniline	0.135	0.333	2.066

obviously, which is an important parameter to evaluate the reaction. It ranges from 0.026 (diethyl phthalate) to 3.327 (basic fuchsin) among 22 organic compounds. The largest reaction rate is more than 127 times larger than that of the smallest one. Fig. 1b shows the reaction rate constants in the condition of neutral solution, the k' value ranges from 0.026 (phloroglucinol) to 2.501 (orange G), the difference of k' value of the two organic compounds is more than 96 times, Fig. 1c shows the reaction rate constants in alkaline condition, the largest reaction rate and the smallest reaction

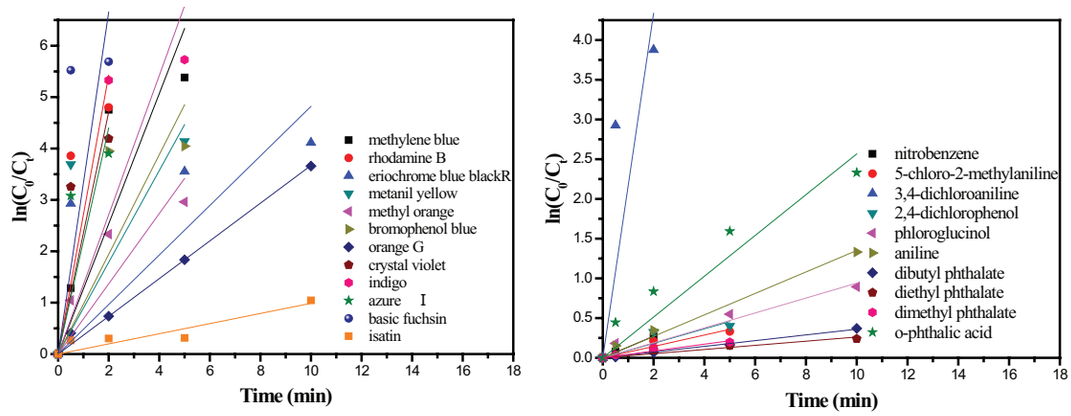

Fig. 1a. The reaction rate constants of colour removal in acid condition.

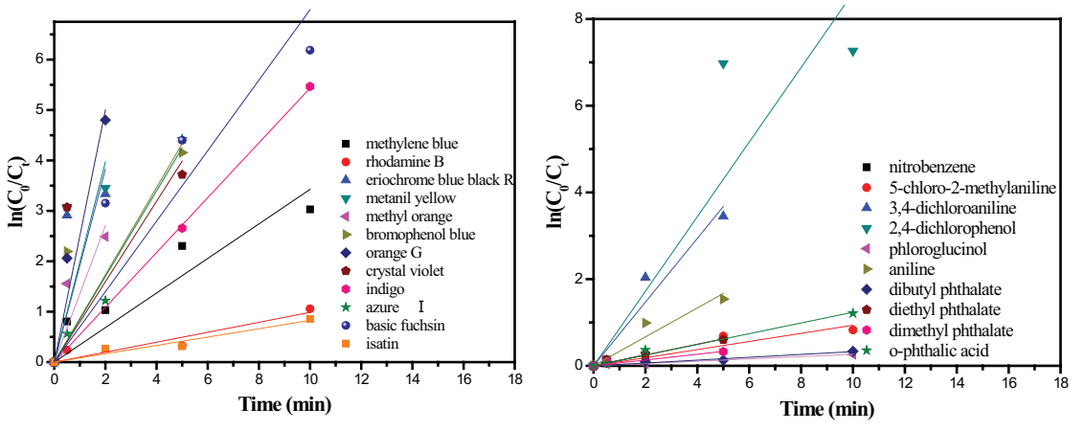

Fig. 1b. The reaction rate constants of colour removal in neutral condition.

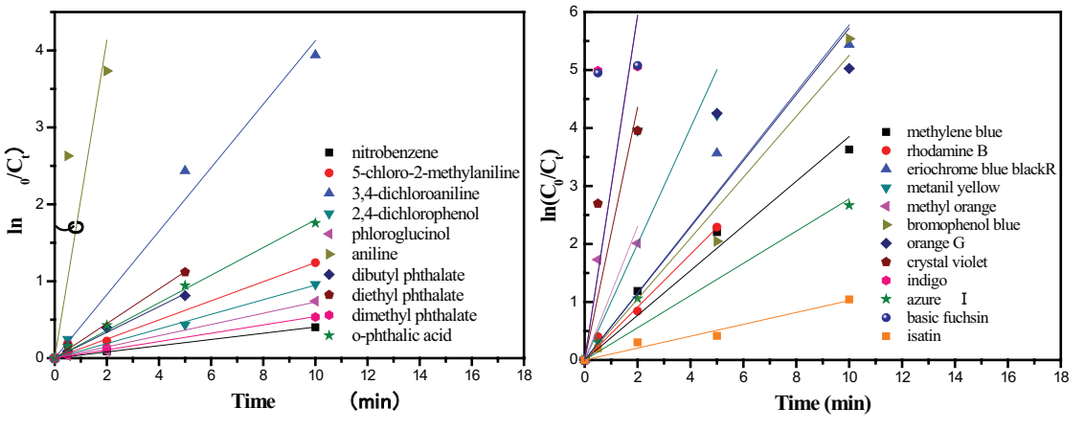

Fig. 1c .The reaction rate constants of colour removal in alkaline condition.

rate constant are 2.972 (basic fuchsin) and 0.040 (nitrobenzene), respectively. The largest reaction rate is more than 74 times larger than that of the smallest one. From the changes of k' value in three conditions, we can easily identify that the selected compounds have huge diversity of their structures, from Table 1, we can also find out each compound has different reaction rate values in different conditions, some organic compounds have huge or small reaction rate in acid condition but have small or huge reaction rate in other two conditions. For example, the reaction rate constant of methylene blue in three conditions are 1.267, 0.342 and 0.385, respectively. On the contrary, the reaction rate constant of orange G in three conditions are 0.367, 2.501 and 0.572, respectively. Since all the compounds we selected are different and diverse, and each compound exhibits different reaction rate in three conditions, it is meaningful to set up a QSAR model with universal applicability to predict the reaction rate of compounds with various structures in acid, neutral and alkaline conditions, respectively. Several studies have been published about QSAR analysis in recent years, among their results, most substances are homologue or have the similar structures. Sudhakaran et al. [16] investigated 22 pharmaceuticals and personal care products (PPCPs) using QSAR method, results showed that only E_{LUMO}–E_{HOMO} had great influence on the degradation of PPCPs. Liu et al. [25] researched 26 kinds of substituted phenols by QSAR analysis, and proposed some basic principles of related chemical reactions. There is a major limitation in application and promotion, and QSAR model in their studies is not appropriate for different compounds with different structures. Consequently, we selected 22 different kinds of organic compounds with various structures to find out some appropriate QSAR models for potassium permanganate oxidative system in different pH conditions.

3.2. Computation results

22 organic compounds and their 17 respective quantum descriptors are listed in Table 2. $q(H^+)$ is obtained by NPA method, which refers to maximum positive partial charge on a hydrogen atom. The value of $q(H^+)$ ranges from 0.217e (methyl orange) to 0.497e (eriochromeblue black R) with the average value of 0.367e. $q(C^-)_x$ and $q(C^-)_n$ are the most negative partial charge on a main-chain carbon atom in the molecule, $q(C^-)_x$ and $q(C^-)_n$ have the average values of 0.293e and –0.330e, respectively. Specifically, the maximum of $q(C^-)_x$ and $q(C^-)_n$ are 0.511e and –0.177e, while the minimum of them are 0.061e and –0.591e. The average of dipole moment (μ) is 5.527 Debye, it should be noticed that the difference between the largest and the smallest value is 14.686 Debye, suggesting the polarity has the huge difference in each organic compound. The larger dipole moment is, the larger polarity is. E_{HOMO} and E_{LUMO} are acronyms for energy of the highest occupied molecular orbital and the lowest unoccupied molecular orbital, respectively. As shown in Table 2, the average of E_{HOMO} and E_{LUMO} are –0.232 eV and –0.058 eV, respectively. E_{HOMO} value ranges from –0.284 eV (methyl orange) to –0.151 eV (crystal violet), while E_{LUMO} value ranges from –0.140 eV (nitrobenzene) to 0.010 eV

Table 2
17 quantum descriptors of 22 kinds of organic compounds

Organic compounds	μ(Debye)	$q(CH^+)_x$(e)	$q(CH^+)_n$(e)	$q(C^-)_x$(e)	$q(C^-)_n$(e)	E_{LUMO}(eV)	E_{HOMO}(eV)	$q(H^+)$(e)
Phloroglucinol	2.686	0.228	0.193	0.376	–0.411	0.010	–0.222	0.462
Dibutyl phthalate	0.791	0.221	0.174	0.423	–0.560	–0.059	–0.266	0.221
Rhodamine B	8.788	0.280	0.190	0.438	–0.581	–0.012	–0.229	0.482
Isatin	4.622	0.219	0.206	0.217	–0.254	–0.105	–0.249	0.409
Methylene blue	12.083	0.239	0.191	0.256	–0.360	–0.127	–0.173	0.239
Eriochromeblue black R	7.110	0.248	0.176	0.453	–0.271	–0.009	–0.276	0.497
Metanil yellow	3.686	0.240	0.199	0.197	–0.248	–0.072	–0.196	0.382
Dimethyl phthalate	1.244	0.238	0.177	0.164	–0.201	–0.073	–0.269	0.238
Bromophenol blue	7.656	0.245	0.212	0.290	–0.257	–0.068	–0.252	0.479
o-phthalic acid	0.077	0.234	0.208	0.153	–0.177	–0.081	–0.275	0.478
Crystal violet	14.763	0.271	0.186	0.261	–0.383	–0.101	–0.151	0.271
3,4-dichloroaniline	5.034	0.219	0.204	0.201	–0.279	–0.027	–0.215	0.381
Basic fuchsin	8.116	0.205	0.166	0.203	–0.591	–0.087	–0.183	0.383
Nitrobenzene	4.541	0.238	0.208	0.061	–0.191	–0.140	–0.276	0.238
Orange G	3.932	0.262	0.196	0.350	–0.265	–0.071	–0.195	0.458
Methyl orange	8.801	0.217	0.136	0.254	–0.287	–0.009	–0.284	0.217
2,4-dichlorophenol	1.140	0.235	0.223	0.306	–0.243	–0.040	–0.243	0.476
Azure i	14.245	0.226	0.166	0.477	–0.296	–0.016	–0.279	0.384
Diethyl phthalate	1.511	0.234	0.180	0.469	–0.584	–0.072	–0.266	0.234
5-chloro-2-methylaniline	3.827	0.216	0.199	0.205	–0.291	–0.010	–0.208	0.383
Indigo	5.235	0.219	0.203	0.511	–0.257	–0.099	–0.204	0.388
Aniline	1.715	0.198	0.193	0.190	–0.266	0.001	–0.198	0.376

Table 2
17 quantum descriptors of 22 kinds of organic compounds

Organic compounds	BO_x	BO_n	$f(+)_x$(e)	$f(+)_n$(e)	$f(-)_x$(e)	$f(-)_n$(e)	$f(0)_x$(e)	$f(0)_n$(e)	E_{B3LYP}(kcal/mol)
Phloroglucinol	1.336	1.310	0.090	0.087	0.098	0.058	0.102	0.074	–458.040
Dibutyl phthalate	1.380	0.960	0.074	0.002	0.044	0.003	0.057	0.003	–924.130
Rhodamine B	1.530	0.962	0.063	–0.005	0.038	–0.007	0.062	–0.006	–1881.430
Isatin	1.392	0.878	0.119	0.026	0.076	0.017	0.096	0.025	–513.200
Methylene blue	1.418	1.038	0.037	0.009	0.037	0.010	0.036	0.012	–1643.400
Eriochrome blue black R	1.532	1.187	0.046	0.001	0.039	0.005	0.046	0.007	–1807.770
Metanil yellow	1.468	1.296	0.048	0.001	0.042	0.004	0.042	0.003	–1645.010
Dimethyl phthalate	0.985	1.377	0.076	0.019	0.045	0.011	0.058	0.020	–688.160
Bromophenol blue	1.405	0.950	0.095	–0.008	0.039	–0.010	0.053	0.002	–11801.150
o-phthalic acid	1.377	0.989	0.078	0.045	0.077	0.011	0.061	0.037	–609.540
Crystal violet	1.503	1.103	0.053	0.002	0.050	0.007	0.051	0.007	–1595.460
3,4-dichloroaniline	1.425	1.272	0.113	0.035	0.081	0.038	0.090	0.050	–1206.920
Basic fuchsin	1.620	1.006	0.073	0.006	0.046	0.008	0.056	0.008	–937.930
Nitrobenzene	1.390	1.323	0.094	0.035	0.060	–0.002	0.077	0.016	–436.860
Orange G	1.514	1.198	0.063	0.010	0.047	0.005	0.055	0.006	–2373.120
Methyl orange	1.459	1.312	0.046	0.015	0.032	0.016	0.038	0.015	–1485.780
2,4-dichlorophenol	1.401	1.259	0.119	0.041	0.079	0.049	0.092	0.060	–1226.790
Azure	1.579	1.045	0.034	0.014	0.033	0.012	0.033	0.012	–1597.480
Diethyl phthalate	1.380	0.983	0.075	0.008	0.044	0.010	0.057	0.009	–766.820
5-chloro-2-methylaniline	1.379	1.003	0.117	0.026	0.116	0.022	0.085	0.024	–786.630
Indigo	1.421	0.964	0.058	0.013	0.050	0.012	0.044	0.016	–875.880
Aniline	1.414	1.288	0.124	0.048	0.125	0.063	0.099	0.092	–287.680

(phloroglucinol). E_{B3LYP} stands for the total energy of a molecule, it has an increasing tendency when the relative molecular mass getting lower. Specifically, the values of E_{B3LYP} for phloroglucinol (molecular masses 126.11) and nitrobenzene (molecular masses 123.11) are much lower than those of other organic molecules.

Bond order (BO_x and BO_n) and fukui indices of 22 organic compounds in this study were calculated by Material Studio 6.1. It can be seen from Table 2, the largest BO_x value is 1.620 (basic fuchsin) while the smallest value is 0.985 (dimethyl phthalate), with the difference of 0.635. And the largest and smallest BO_n values are 1.377 (dimethyl phthalate) and 0.878 (isatin). As for $f(+)_x$, the largest value reaches 0.124e (aniline), while the smallest value is 0.034e (azure 1), the average value is 0.077e. The average values of $f(-)_x$ and $f(0)_x$ are 0.059e and 0.063e, respectively, the largest values of $f(-)_x$ and $f(0)_x$ are 0.125e (aniline) and 0.102e (phloroglucinol), respectively, while the smallest values are 0.032e (methyl orange) and 0.033e (azure I), respectively. The average values of $f(+)_n$, $f(-)_n$ and $f(0)_n$ are 0.020e, 0.016e and 0.022e, respectively. The differences of the largest and smallest $f(+)_n$, $f(-)_n$ and $f(0)_n$ are 0.095e, 0.073e and 0.098e, respectively. From Table 2, it is easy to find out that aniline and phloroglucinol always have high values in the six Fukui indices.

3.3. Construction of QSAR models

22 investigated organic compounds were divided into two groups: training set and test set. As the ln *k′* values in this study were different, and ln *k′* values of the compounds in two sets should cover a range as wide as possible. In view of our previous study , the test set was randomly selected with interval of five [21]. Therefore, we chose four compounds as the test set which have both large and small values in different pH conditions, and others as the training set. The calculated quantum descriptors of 18 compounds in training set and ln *k′* were taken as independent variables and dependent variable, respectively. Consequently, 12 QSAR models were obtained by using MLR analysis and listed in Tables 3–5.

The models along with the associated statistical indices (squared regression coefficient R^2, standard deviations SD, LOO cross-validation (q^2), F value, sig. values and root mean square error (RMSE)) are listed in the Table 3 to Table 5. The optimum equation is testified by comparing R^2 and q^2. In the acid condition, the values of R^2 increase with the number of variables from 0.616 to 0.800. And internal validation q^2 has the same tendency of R^2, with the values ranging from 0.499 to 0.626. In the neutral condition, R^2 of four developed models are 0.279, 0.592, 0.705 and 0.747, respectively. The internal validation q^2 ranges from 0.163 to 0.406. In the alkaline condition, the largest R^2 value and the smallest R^2 value are 0.803 and 0.373, while q^2 values are 0.529 and 0.239, respectively. Both of the largest q^2 values in acid and alkaline conditions are larger than the standard value (0.5). All the F values have the opposite tendency with the increase of variables. The smaller F value is, the more reasonable equation is. The sig. values of all the models are less than 0.05, which indicates the equations are accurate.

Table 3
Regression models for calculating ln*k*′ of organic compounds in acid condition

No.	Model	R^2	SD	q^2	F	Sig.	RMSE
1	$\ln k' = -2.73 + 0.31\ \mu$	0.616	1.063	0.499	25.701	0.000	1.132
2	$\ln k' = 0.65 + 0.29\ \mu + 13.74\ E_{HOMO}$	0.715	0.947	0.482	18.775	0.000	0.898
3	$\ln k' = -1.11 + 0.28\ \mu + 11.97\ E_{HOMO} + 3.79\ q(H^+)$	0.764	0.891	0.562	15.132	0.000	0.794
4	$\ln k' = 0.53 + 0.22\ \mu + 14.26\ E_{HOMO} + 4.89\ q(H^+) - 18.21 f(0)_x$	0.800	0.851	0.626	13.029	0.000	0.724

Table 4
Regression models for calculating ln*k*′ of organic compounds in neutral condition

No.	Model	R^2	SD	q^2	F	Sig.	RMSE
5	$\ln k' = -8.94 + 5.49 BO_x$	0.279	1.212	0.163	12.211	0.008	1.469
6	$\ln k' = -9.83 + 7.52 BO_x + 5.84 q(C^-)_n$	0.592	0.942	0.202	11.128	0.001	0.888
7	$\ln k' = -7.09 + 6.57 BO_x + 5.91 q(C^-)_n - 21.42 f(+)_n$	0.705	0.829	0.208	10.861	0.000	0.699
8	$\mathrm{lln}\ k' = -3.86 + 5.84 BO_x + 6.13 q(C^-)_n - 24.88 f(+)_n + 8.01 E_{HOMO}$	0.747	0.795	0.406	9.620	0.000	0.681

Table 5
Regression models for calculating ln*k*′ of organic compounds in alkaline condition

No.	Model	R^2	SD	q^2	F	sig.	RMSE
9	$\ln k' = 4.36 + 22.34 E_{HOMO}$	0.373	1.155	0.239	14.530	0.007	1.335
10	$\ln k' = 7.24 + 25.38 E_{HOMO} - 33.60 f(0)_x$	0.653	0.887	0.446	14.129	0.000	0.788
11	$\ln k' = 12.84 + 26.39 E_{HOMO} - 39.02 f(0)_x - 21.60 q(CH^+)_x$	0.744	0.789	0.518	13.572	0.000	0.623
12	$\ln k' = 13.62 + 27.88 E_{HOMO} - 41.19 f(0)_x - 20.96 q(CH^+)_x + 8.10\ E_{LUMO}$	0.803	0.710	0.529	13.228	0.000	0.517

According to previous studies of our team, a good QSAR model usually had three or four independent variables [26]. Besides, the equations with four variables have the minimum SD values and RMSE values in three pH conditions, which demonstrated the equations we selected have the smallest deviation and dispersion and high statistical significance. Thus, the equations with four variables were selected as the optimum equations. Based on the evaluation criteria above, we can consider the model 4, model 8 and model 12 in the following tables are reasonable although the q^2 value of model 8 cannot satisfy the internal evaluation criteria. In addition, the difference between predicted values and experimental values for reaction rate constants of the most suitable models in three conditions can be seen from Fig. 2a, Fig. 2b and Fig. 2c, respectively.

3.4. Analysis and discussion of derived models

For the acid condition, the derived model contains four variables μ, E_{HOMO}, $q(H^+)$ and $f(0)_x$. Each descriptor plays an important role in the degradation process. According to the equation of model 4, ln*k*′ increases with μ, E_{HOMO} and $q(H^+)$, the larger the former three descriptors are, the larger ln *k*′ will be. And it is easier for 18 compounds to be degraded with larger μ, E_{HOMO} and $q(H^+)$ values, and vice versa. E_{HOMO} is one of the frontier molecular orbitals, which is widely used in developing QSAR models for studying the various mechanisms of organic compounds [27]. Compounds with high E_{HOMO} values indicating that the electrons of molecular orbital are instability, and easier to lose. The descriptor $q(H^+)$ reflects the characteristic of the charge distribution of a molecule. It is related to the maximum positive partial charge on a hydrogen atom in the molecule that is usually connected to the electron withdrawing [28]. Furthermore, ln *k*′ decreases with $f(0)_x$. When $f(0)_x$ is larger, it is harder for C–H bonds of aliphatic hydrocarbons and N–H bonds of amines to be ruptured. Compounds with high $f(0)_x$ values have strong tolerance to be oxidized because they have a higher barrier to be attacked. Hence, when the $f(0)_x$ values of organic compounds are larger, they are harder to be attacked and then degraded slowly.

In the neutral condition, the selected model also consists of four variables (BO_x, $q(C^-)_n$, $f(+)_n$ and E_{HOMO}), BO_x means the maximum of chemical bonds between a pair of atoms in a molecule, we can contrast Table 1 and Table 2, the larger BO_x value is, the larger *k*′ value will be, expect basic fuchsin, the BO_x value of basic fuchsin is 1.620, while the *k*′ value is 0.699. $q(C^-)_n$ means characteristic of the charge distribution of C–C bond and has a bearing on the most negatively charged atom in a molecule. When the $q(C^-)_n$ is lower, the carbon atom has more negative values than that of other carbon atoms and degraded slowly. For instance, basic fuchsin has the smallest $q(C^-)_n$ value, leading to its low reaction rate(0.066). E_{HOMO} is positively correlated with ln *k*′, while $f(+)_n$ is negatively correlated with ln *k*′.

In the alkaline condition. The optimum model includes four variables, E_{HOMO}, $f(0)_x$, $q(CH^+)_x$ and E_{LUMO}. From model

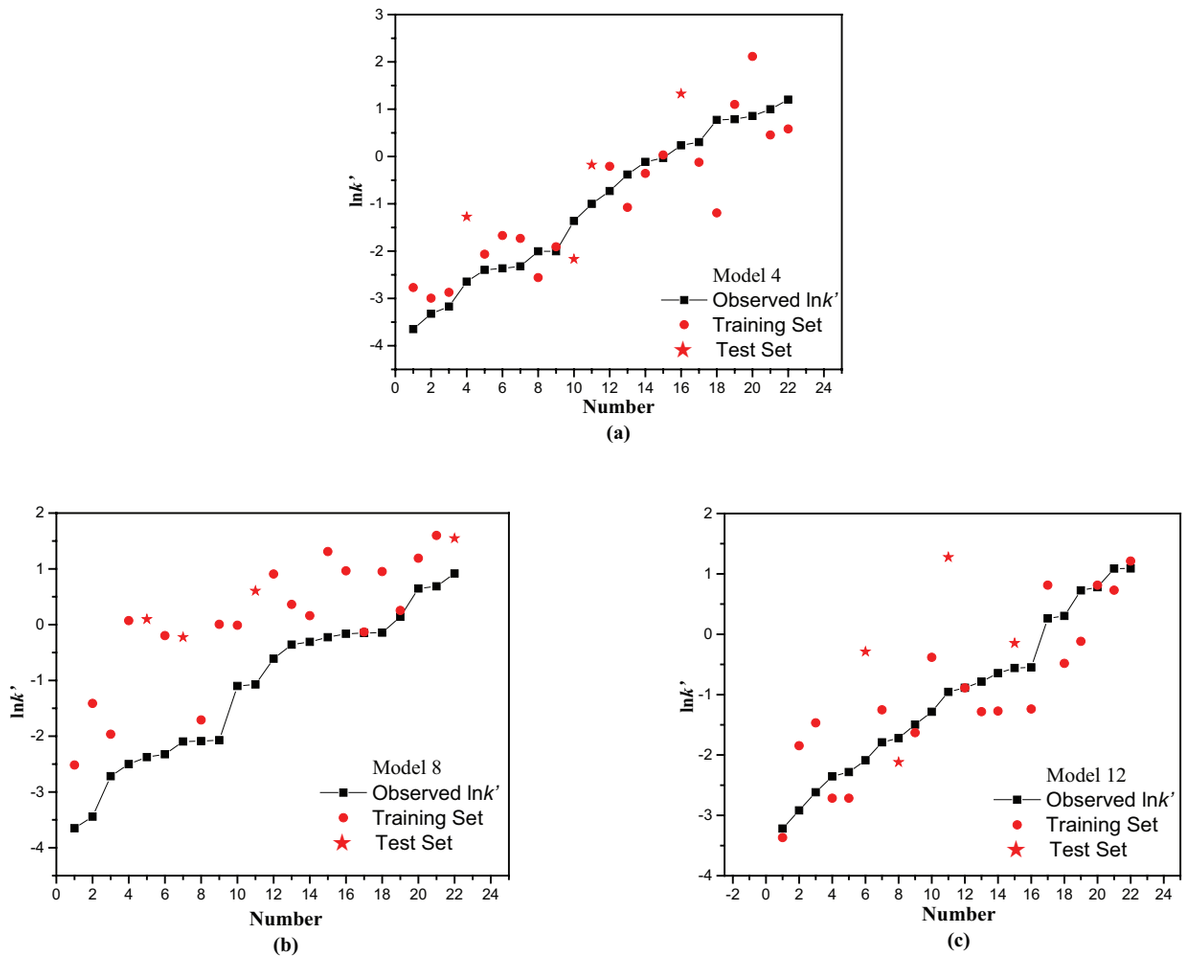

Fig. 2. Predicted values vs. experimental values of ln k' for selected models.

12, we can discover that ln k' increases with E_{HOMO} and E_{LUMO}, while decreases with $f(0)_x$ and $q(CH^+)_x$. It means that there is a negative correlation relationship between ln k' and $q(CH^+)_x$. For example, the calculated $q(CH^+)_x$ value of aniline is the lowest (0.198e), resulted the fast reaction rate with k' value of 2.066.

Compared with the optimum models we selected, all of the three models contain E_{HOMO} and Fukui indices, indicating that the energy of the highest occupied molecular orbital and the electronic layout are the most important factors in potassium permanganate degradation process, even though there is a little difference, they are all a part of Fukui indices. $f(0)_x$ appears in both acid condition and alkaline condition, indicating that the whole radical was attacked by potassium permanganate in acid and alkaline conditions, while $f(+)_n$ influences the reaction rate in the neutral condition, suggesting that nucleophilic attack is preferred in neutral condition. We can also see from the equations that all of the Fukui indices have negative correlation with the reaction rate and E_{HOMO} has positive correlation with the reaction rate. Besides, Table 6 to Table 8 show the pairwise correlation coefficients of lnk' and independent variables, and the correlation coefficients of each independent

Table 6
Correlation coefficient matrix for variables of model 4

	ln k'	μ	E_{HOMO}	$q(H^+)$	$f(0)_x$
ln k'	1.000				
μ	0.761	1.000			
E_{HOMO}	0.419	0.292	1.000		
$q(H^+)$	0.233	−0.106	0.054	1.000	
$f(0)_x$	−0.466	−0.530	0.052	0.273	1.000

Table 7
Correlation coefficient matrix for variables of model 8

	ln k'	BO_x	$q(C^-)_n$	$f(+)_n$	E_{HOMO}
ln k'	1.000				
BO_x	0.553	1.000			
$q(C^-)_n$	0.291	−0.316	1.000		
$f(+)_n$	−0.416	−0.318	0.285	1.000	
E_{HOMO}	0.207	0.280	−0.183	−0.063	1.000

Table 8
Correlation coefficient matrix for variables of model 12

	ln k'	E_{HOMO}	$f(0)_x$	$q(CH^+)_x$	E_{LUMO}
ln k'	1.000				
E_{HOMO}	0.537	1.000			
$f(0)_x$	–0.431	–0.052	1.000		
$q(CH^+)_x$	0.052	0.099	–0.312	1.000	
E_{LUMO}	0.038	–0.188	0.261	–0.159	1.000

variable. From the tables, we can find that the correlation coefficients order between ln k' and independent variables in three conditions are different, for the acid condition, the order is μ (0.761)> $f(0)_x$ (–0.466) > E_{HOMO} (0.419)> $q(H^+)$ (0.233), the order in neutral condition is BO_x (0.553)> $f(+)_n$ (–0.416)>$q(C^-)_n$ (0.291)>E_{HOMO} (0.207), the order in alkaline is as follows: E_{HOMO} (0.537)> $f(0)_x$ (–0.431)>$q(CH^+)_x$ (0.052)>E_{LUMO} (0.038). The most important factors in three conditions are μ, BO_x and E_{HOMO}, respectively. Generally, electric charge also influences the degradation in three conditions, specifically, the positive partial charge on hydrogen atom affects the degradation in acid condition, for the neutral condition the negative partial charge on a carbon atom is important, while the positive partial charge on a hydrogen atom linked with a carbon atom influences the degradation in alkaline condition.

3.5. Validation of models

To test the significance and stability of each variable in each model, the standard regression coefficients and t test were carried out by SPSS 20.0. The results are displayed in Table 9. All the absolute t values are larger than the standard one (1.740) except the t value of E_{HOMO} in model 8 and t value of $q(CH^+)_x$ in model 12. All the sig. values are smaller than 0.05. Moreover, multicollinearity between the variables of each model was checked by calculating their variation inflation factors (VIF) to evaluate the correlation degree of each independent variable in the models [29]. The equation of VIF is $VIF=1/(1–r^2)$, in which r is the correlation coefficient of multiple regression between one variable and the others in the equation. If VIF = 0, there is no inter correlation in each variable; If VIF ranges from 1.0 to 5.0, the relevant equation is acceptable; If VIF is larger than 10.0, the regression equation is unstable and need to recheck. According to Table 9, all VIF values of variables are larger than 1.0 but far less than 5.0, the max-VIF values for each model are 1.623, 1.260 and 1.188, respectively, which suggest that the variables of models do not have multicollinearity.

Y- Randomization test was also used to examine whether the models are robust and statistical or not. The ln k' values were randomly shuffled and as the dependent variable, a new model was developed using original independent variable matrix. After ten times of repetition, the derived R^2 and q^2 values are expected to be lower than those of the original models. As shown in Table 10, all of the R^2 and q^2 values are much smaller than the original values. From the results, we can consider that the original models in this study have no possibility of chance correlation and show good robustness.

Table 9
Checking statistical values for model 4, model 8 and model 12

		Regression coefficients	t	sig.	VIF
Model 4	μ	0.22 ± 0.991	3.847	0.001	1.623
	E_{HOMO}	14.26 ± 0.039	1.861	0.008	1.170
	$q(H^+)$	4.89 ± 0.100	2.837	0.011	1.084
	$f(0)_x$	–18.21 ± 0.023	–1.750	0.030	1.581
Model 8	BO_x	5.84 ± 0.573	3.867	0.001	1.260
	$q(C^-)_n$	6.13 ± 0.617	4.316	0.000	1.174
	$f(+)_n$	24.88 ± 0.398	–2.795	0.012	1.164
	E_{HOMO}	8.01 ± 0.175	1.263	0.024	1.100
Model 12	E_{HOMO}	27.88 ± 0.648	4.315	0.000	1.060
	$f(0)_x$	–41.19 ± 0.620	–3.902	0.001	1.188
	$q(CH^+)_x$	–20.96 ± 0.266	–1.714	0.005	1.129
	E_{LUMO}	8.10 ± 0.278	–1.800	0.009	1.126

Table 10
Y-Randomization results of three models

Iteration	Model 4		Model 8		Model 12	
	R^2	q^2	R^2	q^2	R^2	q^2
Original	0.800	0.626	0.747	0.406	0.803	0.529
1	0.017	0.135	0.372	0.077	0.091	0.263
2	0.056	0.317	0.248	0.016	0.095	0.230
3	0.362	0.013	0.438	0.131	0.360	0.027
4	0.263	0.005	0.278	0.012	0.437	0.105
5	0.079	0.281	0.018	0.280	0.156	0.171
6	0.099	0.176	0.060	0.308	0.104	0.396
7	0.092	0.223	0.065	0.339	0.086	0.385
8	0.129	0.139	0.127	0.046	0.112	0.289
9	0.130	0.154	0.115	0.062	0.146	0.156
10	0.076	0.295	0.128	0.057	0.080	0.218

Besides, four organic compounds were selected as the test set, including methylene blue, o-phthalic acid, orange G and 5-chloro-2-methylaniline. External validation (Q^2_{EXT}) was employed for evaluating the predictive potential of the developed models. Research [30] showed that if the values of Q^2_{EXT} were larger than 0.5, the model can be considered as predictive, robust and acceptable. The Q^2_{EXT} values of the three models are 0.997, 0.314 and 0.264, respectively.

Another metric r^2_m was calculated to validate the prediction ability of the models as well. Roy et al. [31] suggested that we can use the stricter statistical parameters (r^2_m, r^2_m, r'^2_m and $\overline{r^2_m}$) to avoid over estimation of the models. If Δr^2_m <0.2, r^2_m, r'^2_m, and $\overline{r^2_m}$ are all larger than 0.5, the models we developed have good prediction, otherwise the models cannot be accepted [32]. The results are showed in Table 11, we can clearly see that only model 4 satisfy the verification.

Therefore, from the validations we mentioned above, we can cognizance that only the model in acid condition can be considered as qualified.

Table 11
Prediction ability validation results of three models

Statistic	r_m^2	r'^2_m	Δr_m^2	$\overline{r_m^2}$
Model 4	0.5463	0.5799	0.0038	0.5631
Model 8	0.0442	0.9377	0.8935	0.4909
Model 12	0.1804	0.2338	0.0053	0.1036

4. Conclusions

Based on the experimental data, reaction rate constants of 22 organic compounds were calculated in acid, neutral and alkaline conditions, the reaction rate was determined as the first-order reaction, QSAR models for potassium permanganate system were constructed by 17 quantum descriptors of 22 organic compounds which were calculated by Gaussian 09 and Material Studio 6.1 using multiple linear regression analysis. μ, $E_{HOMO,}$ E_{LUMO}, $q(H^+)$, $q(C^-)_n$, $q(CH^+)_x$, BO_x and Fukui indices were important factors in QSAR models in different conditions. E_{HOMO} and Fukui indices appeared in all the selected models which indicated that the two factors were the most important properties in potassium permanganate oxidation process. However, among the 12 models, model 4 which was developed in acid condition satisfied all the criteria with R^2=0.800, q^2=0.626, F=13.029, SD=0.851, Q^2_{EXT}=0.997. The results of t test, VIF test, Y-Randomization validation and r_m^2 validation also suggested that model 4 exhibited good robust and predictive. The equation of model 4 is $\ln k'=0.53+0.22\mu+14.26E_{HOMO}+4.89q(H^+)-18.21f(0)_x$. Among the four quantum descriptors in model 4, μ is the most significant influence on the reaction rate constants, following by E_{HOMO}, $f(0)_x$, $q(H^+)$. Furthermore, according to the models in different conditions, the recommended model provides some suggestions that we can get better results in acid condition for the degradation of organic compounds in potassium permanganate system and using the proposed model, we can estimate the reaction rate constants with less time and effort.

Acknowledgements

This work was supported by the National Science Foundation of China (Project No. NSFC 21537002), and the program for New Century Excellent Talents in Shanghai Jiao Tong University.

Reference

[1] Z. Aksu, Application of biosorption for the removal of organic pollutants: a review, Process Biochem., 40 (2005) 997–1026.
[2] A. Bhatnagar, W. Hogland, M. Marques, M. Sillanpää, An overview of the modification methods of activated carbon for its water treatment applications, Chem. Eng. J., 219 (2013) 499–511.
[3] G. Lofrano, S. Meric, G.E. Zengin, D. Orhon, Chemical and biological treatment technologies for leather tannery chemicals and wastewaters: a review, Sci. Total Environ., 461–462 (2013) 265–281.
[4] B.D. Blair, J.P. Crago, C.J. Hedman, R.J. Treguer, C. Magruder, L.S. Royer, R.D. Klaper, Evaluation of a model for the removal of pharmaceuticals, personal care products, and hormones from wastewater, Sci. Total Environ., 444 (2013) 515–521.
[5] S.W. Krasner, W.A. Mitch, D.L. McCurry, D. Hanigan, P. Westerhoff, Formation, precursors, control, and occurrence of nitrosamines in drinking water: a review, Water Res., 47 (2013) 4433–4450.
[6] S. Dash, S. Patel, B.K. Mishra, Oxidation by permanganate: synthetic and mechanistic aspects, Tetrahedron, 65 (2009) 707–739.
[7] Y.E. Yan, F.W. Schwartz, Oxidative degradation and kinetics of chlorinated ethylenes by potassium permanganate, J. Contam. Hydrol., 37 (1999) 343–365.
[8] M. Schnarr, C. Truax, G. Farquhar, E. Hood, B. Stickney, Laboratory and controlled field experiments using potassium permanganate to remediate trichlorothylene and perchloroethylene DNAPLs in porous media, J. Contam. Hydrol., 19 (1998) 205–224.
[9] S.Y. Pang, J. Jiang, Y. Gao, Y. Zhou, X. Huangfu, Y. Liu, J. Ma, Oxidation of flame retardant tetrabromobisphenol a by aqueous permanganate: reaction kinetics, brominated products, and pathways, Environ. Sci. Technol., 48 (2014) 615–623.
[10] H.W. Rachel, G.T. Paul, Kinetics of contaminant degradation by permanganate, Environ. Sci. Technol., 40(3) (2006) 1055–1061.
[11] G. Rupper, R. Bauer, G. Heisler, S. Novalic, Mineralization of cyclic organic water contaminants by the photo-Fenton reaction-influence of structure and substituents, Chemosphere, 27 (1993) 1339–1347.
[12] Y. Ren, H. Liu, X. Yao, M. Liu, Prediction of ozone tropospheric degradation rate constants by projection pursuit regression, Anal. Chim. Acta, 589 (2007) 150–158.
[13] S. Bruzzone, C. Chiappe, S.E. Focardi, C. Pretti, M. Renzi, Theoretical descriptor for the correlation of aquatic toxicity of ionic liquids by quantitative structure–toxicity relationships, Chem. Eng. J., 175 (2011) 17–23.
[14] A.M. Redding, F.S. Cannon, S.A. Snyder, B.J. Vanderford, A QSAR-like analysis of the adsorption of endocrine disrupting compounds, pharmaceuticals, and personal care products on modified activated carbons, Water Res., 43 (2009) 3849–3861.
[15] A.A. Toropov, A.P. Toropova, T. Puzyn, E. Benfenati, G. Gini, D. Leszczynska, J. Leszczynski, QSAR as a random event: modeling of nanoparticles uptake in PaCa2 cancer cells, Chemosphere, 92 (2013) 31–37.
[16] S. Sudhakaran, J. Calvin, G.L. Amy, QSAR models for the removal of organic micropollutants in four different river water matrices, Chemosphere, 87 (2012) 144–150.
[17] C. Cardenas, N. Rabi, P.W. Ayers, C. Morell, P. Jaramillo, P. Fuentealba, Chemical reactivity descriptors for ambiphilic reagents: dual descriptor, local hypersoftness, and electrostatic potential, J. Phys. Chem. A., 113 (2009) 8660–8667.
[18] S.B. Liu, Conceptual density functional theory and some recent developments, Acta Phys.-Chim. Sin., 25(3) (2009) 590–600.
[19] X. Li, W. Zhao, J. Li, J. Jiang, J. Chen, J. Chen, Development of a model for predicting reaction rate constants of organic chemicals with ozone at different temperatures, Chemosphere, 92 (2013) 1029–1034.
[20] H. Zhu, W. Guo, Z. Shen, Q. Tang, W. Ji, L. Jia, QSAR models for degradation of organic pollutants in ozonation process under acidic condition, Chemosphere, 119 (2015) 65–71.
[21] L. Jia, Z. Shen, W. Guo, Y. Zhang, H. Zhu, W. Ji, M. Fan, QSAR models for oxidative degradation of organic pollutants in the Fenton process, J. Taiwan Inst. Chem. E., 46 (2015) 140–147.
[22] P. Su, H. Zhu, Z. Shen, QSAR models for removal rates of organic pollutants adsorbed by in situ formed manganese dioxide under acid condition, Environ. Sci. Pollut. Res., 23 (2016) 3609–3620.
[23] R.G. Parr, W. Yang, Density functional approach to the FrontieR-Electron Theory of Chemical Reactivity, J. Am. Chem. Soc., 106 (1984) 4049–4050.
[24] F.D. Proft, C.V. Alsenoy, A. Peeters, W. Langenaeker, P. Geerlings, Atomic charges, dipole moments, and Fukui functions using the Hirshfeld partitioning of the electron density, J. Comput. Chem., 23 (2002) 1189–1209.
[25] H. Liu, J. Tan, H.X. Yu, L.S. Wang, Z.Y. Wang, Determination of the appareant rate constants for ozone degradation of substituted phenols and QSPR/QSAR analysis, Int. J. Environ. Res., 4(3) (2010) 507–512.

[26] H. Zhu, Z. Shen, Q. Tang, W. Ji, L. Jia, Degradation mechanism study of organic pollutants in ozonation process by QSAR analysis, Chem. Eng. J., 255 (2014) 431–436.
[27] M.A. Turabekova, B.F. Rasulev, M.G. Levkovich, N.D. Abdullaev, J. Leszczynski, *Aconitum* and *Delphinium* sp. alkaloids as antagonist modulators of voltage-gated Na+ channels. AM1/DFT electronic structure investigations and QSAR studies, Comput. Biol. Chem., 32 (2008) 88–101.
[28] J.L. Jiang, X.A. Yue, Q.F. Chen, Z. Gao, Determination of ozonization reaction rate constants of aromatic pollutants and QSAR study, Bull. Environ. Contam. Toxicol., 85 (2010) 568–572.
[29] G.R. Famini, C.A. Penski, L.Y. Wilson, Using theoretical descriptors in quantitative structure activity relationships: some physicochemical properties, J. Phys. Org. Chem., 5 (1992) 395–408.
[30] B. Ma, H. Chen, M. Xu, T. Hayat, Y. He, J. Xu, Quantitative structure-activity relationship (QSAR) models for polycyclic aromatic hydrocarbons (PAHs) dissipation in rhizosphere based on molecular structure and effect size, Environ. Pollut., 158 (2010) 2773–2777.
[31] K. Roy, I. Mitra, S. Kar, P.K. Ojha, R.N. Das, H. Kabir, Comparative studies on some metrics for external validation of QSPR models, J. Chem. Inf. Model, 52 (2012) 396–408.
[32] P.K. Ojha, I. Mitra, R.N. Das, K. Roy, Further exploring r^2_m metrics for validation of QSPR models, Chemometr. Intell. Lab., 107 (2011) 194–205.