

PREDICTION OF REACTION RATE CONSTANTS OF HYDROXYL RADICAL WITH ORGANIC COMPOUNDS

ZHEN CHEN¹, XINLIANG YU^{*1}, XIANWEI HUANG¹ AND SHIHUA ZHANG^{*1,2}¹ College of Chemistry and Chemical Engineering, Hunan Institute of Engineering, Xiangtan, Hunan 411104, China² Network Information Center, Hunan Institute of Engineering, Xiangtan, Hunan 411104, China

(Received: June 21, 2012 - Accepted: January 21, 2013)

ABSTRACT

In this work we have performed a QSAR study of atmospheric reactions with hydroxyl radical, at the B3LYP level of theory with 6-31G(d) basis set. Molecular descriptors selected by applying multiple linear stepwise regression (MLR) analysis were used to predict the reaction rate constants ($-\log k_{\text{OH}}$) of OH radicals with organic compounds in the atmosphere, including 98 alkenes and 80 aromatic compounds. For setting our reactions, we have calculated 98 alkenes and 80 aromatic compounds. A four-descriptor MLR model ($r_{\text{ms}} = 0.102$ and $R^2 = 0.938$) for 98 alkenes was developed based on the number of R=CHX functional group counts, E_{HOMO} parameter, and Fukui indices of the double-bonded C atoms. We found vinyl chemicals with conjugated double bonds and electron-donor substituents are the most reactive systems; while alkenes with multiple halogen substitutions are the least reactive molecules. Additionally, a three-descriptor MLR model ($r_{\text{ms}} = 0.282$ and $R^2 = 0.910$) was built to predict OH radical rate constants for 80 aromatic compounds, which was dominated by the E_{HOMO} parameter, a topological descriptor for steric hindrance, and the most positive net atomic charge on hydrogen atoms. Aromatics with electron-donor and electron-acceptor groups, respectively, possess high and low degradation rates. The halogen aromatics are less reactive, especially for aromatics with multiple halogen substitutions. In comparison to existing models, the two models obtained in this paper show better statistical quality.

Keywords: alkenes; aromatics; density functional theory; quantum chemical; structure–activity relationships; rate constant.

1. INTRODUCTION

In recent decades, large amounts of organic compounds have emitted into the atmosphere. They may cause photochemical air pollution, acid deposition, long-range transport of chemicals, changes of the stratospheric ozone layer, global weather modification, etc., through a complex array of chemical and physical transformations.¹ These chemicals can be chemically transformed in the troposphere by reactions with photochemically generated oxidants like OH radicals and ozone during the daytime and NO₃ radicals at night.²⁻⁴

The reactions of organic pollutants with ·OH in the atmosphere is of great concern because it is the primary process for their degradation and transformation in the daytime, and the behavior and fate of organic compounds in the atmosphere should be assessed according to their reaction rate constants. But only a limited number of experimentally measured rate constants are available. Moreover, all the experimental methods are time-consuming, laborious, costly and equipment dependent.⁵ Reaction rates can be predicted with quantitative structure–activity relationship (QSAR) models.⁶ QSAR is the process by which molecular structural descriptors are quantitatively correlated with chemical properties or activities of molecules. Being cost-effective and rapid estimation methods, several QSAR models have been successfully developed for predicting the rate constants of many structural heterogeneous compounds with ·OH radicals,⁷⁻¹⁵ NO₃ radicals and O₃.¹⁵⁻¹⁸

Bakken and Jurs developed five-descriptor QSAR models for k_{OH} of 57 unsaturated hydrocarbons and ten-descriptor QSAR models for 312 compounds with multiple linear regression (MLR) and artificial neural networks (ANNs).⁷ Pompe *et al.* used topological indices to develop a 6-parameter MLR model ($r_{\text{ms}} = 0.115$) for k_{OH} of 58 unsaturated organic compounds.⁸ Gramatica *et al.* constructed three MLR models for k_{OH} with 4–6 predictor variables.^{9,10}

The prediction r_{ms} errors were not less than 0.4 log units. Öberg developed a QSAR model with partial least squares (PLS) regression. The model had the prediction standard error of 0.501 log units, through selecting 333 descriptors and compressing to 7 latent variables.¹¹ Recently, Fatemi and Baher successfully developed six QSAR models for k_{OH} of 98 alkenes, by applying several chemometric tools including MLR, genetic algorithms (GAs), ANNs and support vector machines (SVMs).¹² Wang *et al.* used 22 molecular descriptors to build a PLS model having a r_{ms} error of 0.430 for the test set.¹³

All the models stated above are based on molecular structures that were optimized with semi-empirical quantum chemistry methods, AM1 or PM3. In addition, QSAR models based on ANNs or SVMs possess complicated structures, which are unfavorable to extension and application of the models. The aim of this paper is to produce new robust QSAR models for k_{OH} of 98 alkenes and 80 aromatics, by applying the density functional theory (DFT) for geometry optimization.

2. COMPUTATIONAL DETAILS

2.1. Data collection

The experimental rate constants (k_{OH}) for the reactions of OH radical with 98 alkenes¹² and 80 aromatics¹³ are listed in Tables 1 and 2 respectively. These alkenes have been studied by Fatemi and Baher.¹² Experimental values were measured at 25°C and 101.3 kPa and reported in cm³ s⁻¹ molecule⁻¹. To obtain positive values for k_{OH} , the listed numerical values were the negative logarithm of experimental k_{OH} values reported. The two entire sets contained a wide range of rate constant values, and were characterized by a high degree of structural variety. The experimental $-\log k_{\text{OH}}$ data in each table were split into a training set (50%) and a test set (50%). The training set was used to develop a MLR model, which was validated with the test set.

Table 1. Molecular descriptors and $-\log k_{\text{OH}}$ values for 98 alkenes.

No.	Name	f_{1C}	f'_{2C}	N_X	E_{HOMO}	Exp.	Pred.
Training set							
1	α -Phellandrene	-0.105505	-0.011451	0	-0.201328	9.50	9.74
2	trans-Ocimene	-0.057367	-0.083843	0	-0.214890	9.60	9.79
3	Terpinolene	-0.038062	-0.025215	0	-0.215923	9.65	9.72
4	1-3-5-Hexatriene	-0.099162	-0.081043	0	-0.209238	9.66	9.84
5	2-5-Dimethyl-2-4-hexadiene	-0.039113	-0.050633	0	-0.194584	9.68	9.51
6	γ -Terpinene	-0.067472	-0.029318	0	-0.217922	9.75	9.82

7	1-3-Cyclohexadiene	-0.089259	-0.158518	0	-0.205526	9.79	9.82
8	cis-2-trans-4-Hexadiene	-0.100778	-0.127656	0	-0.207100	9.81	9.85
9	1-3-Cycloheptadiene	-0.037108	-0.198453	0	-0.216948	9.86	9.82
10	trans-trans-2-4-Hexadiene	-0.094374	-0.124102	0	-0.205513	9.87	9.81
11	2-3-Dimethyl-1-3-butadiene	-0.138670	0.000397	0	-0.224837	9.91	10.07
12	β -Caryophyllene	-0.053939	-0.023733	0	-0.219970	9.92	9.81
13	1-5-Dimethyl-1-5-hexadiene	-0.092121	-0.008002	0	-0.233258	9.92	10.04
14	trans-1-3-Hexadiene	-0.119738	-0.109274	0	-0.215871	9.95	9.98
15	2-3-Dimethyl-2-butene	-0.070417	-0.070417	0	-0.217816	9.96	9.85
16	Dimethylketene	-0.115355	-0.184448	0	-0.211921	9.97	9.97
17	2-Methyl-1-3-butadiene	-0.144574	-0.000815	0	-0.226117	10.00	10.10
18	1-4-Cyclohexadiene	-0.053513	-0.162187	0	-0.226180	10.00	9.94
19	trans-1-3-Pentadiene	-0.122774	-0.114538	0	-0.216601	10.00	10.00
20	2-3-Dimethyl-2-pentene	-0.079994	-0.081041	0	-0.211185	10.01	9.81
21	1-Methylcyclohexene	-0.127708	-0.057063	0	-0.224366	10.03	10.07
22	2-Methyl-2-pentene	-0.126040	-0.060845	0	-0.221978	10.04	10.04
23	trans-1-4-Hexadiene	-0.074911	-0.100508	0	-0.230689	10.04	10.01
24	2-Methyl-2-butene	-0.123392	-0.051252	0	-0.225635	10.06	10.07
25	2-Heptene	-0.109973	-0.237771	0	-0.230412	10.07	10.17
26	2-Methyl-1-4-pentadiene	-0.074836	-0.090135	0	-0.239354	10.10	10.10
27	Cycloheptene	-0.103264	-0.231826	0	-0.232142	10.13	10.17
28	cis-4-Octene	-0.114484	-0.243781	0	-0.231550	10.14	10.20
29	trans-4-Octene	-0.109421	-0.235237	0	-0.229560	10.16	10.16
30	Cyclohexene	-0.112426	-0.248897	0	-0.233459	10.17	10.21
31	trans-2-Pentene	-0.110031	-0.237968	0	-0.230681	10.17	10.17
32	cis-2-Pentene	-0.115148	-0.245782	0	-0.232569	10.18	10.21
33	4-Methyl-1-cyclohexene	-0.112632	-0.247476	0	-0.233595	10.21	10.22
34	2-Methyl-1-butene	-0.175620	-0.037029	0	-0.239266	10.22	10.34
35	trans-4-Methyl-2-pentene	-0.105068	-0.228773	0	-0.235456	10.22	10.21
36	Sabinene	-0.143334	-0.005605	0	-0.221302	10.25	10.05
37	α -Pinene	-0.120493	-0.040878	0	-0.218341	10.26	9.98
38	trans-4,4-Dimethyl-2-pentene	-0.126437	-0.227621	0	-0.232897	10.26	10.24
39	1,4-Pentadiene	-0.091577	-0.130875	0	-0.240502	10.27	10.18
40	2,3,3-Trimethylbutene	-0.174522	-0.038531	0	-0.239358	10.30	10.34
41	Longifolene	-0.146700	-0.018208	0	-0.229710	10.35	10.15
42	1-Heptene	-0.140608	-0.211285	0	-0.246201	10.39	10.40
43	1-Octene	-0.130345	-0.197227	0	-0.246141	10.40	10.37
44	1-Hexene	-0.155653	-0.233817	0	-0.246307	10.43	10.46
45	3-Methyl-1-butene	-0.157667	-0.221693	0	-0.249824	10.49	10.49
46	1-Pentene	-0.154580	-0.233360	0	-0.246525	10.50	10.46
47	Ketene	-0.262671	-0.177630	0	-0.240468	10.76	10.65
48	1-Bromoethene	-0.127509	-0.135847	1	-0.254441	11.17	11.17
49	trans-1,2-Difluoroethene	-0.173960	-0.332287	2	-0.253777	12.13	12.13
	Test set						
50	α -Terpinene	-0.052336	-0.054852	0	-0.192386	9.44	9.53
51	Ocimene	-0.057367	-0.083843	0	-0.214890	9.60	9.79
52	α -Humulene	-0.042641	-0.001900	0	-0.216163	9.65	9.73

53	Myrcene	-0.078040	-0.005205	0	-0.222559	9.67	9.89
54	3-7-Dimethyl-1-6-octadiene	-0.051451	-0.063437	0	-0.223194	9.74	9.85
55	β -Phellandrene	-0.131298	0.003156	0	-0.213643	9.78	9.93
56	2-4-Dimethyl-1-3-butadiene	-0.133652	0.007906	0	-0.214725	9.80	9.95
57	Limonene	-0.075281	-0.003234	0	-0.225634	9.84	9.91
58	3-Methyl-1-3-pentadiene	-0.118096	-0.101050	0	-0.213908	9.87	9.95
59	4-Methyl-1-3-pentadiene	-0.112913	-0.100647	0	-0.207910	9.88	9.88
60	trans-3-Methyl-2-pentene	-0.123210	-0.064764	0	-0.217403	9.91	9.99
61	2-5-Norbornadiene	-0.065542	-0.182030	0	-0.216970	9.92	9.88
62	1-2-Dimethylcyclohexene	-0.077024	-0.077025	0	-0.216046	9.93	9.85
63	trans-1-3-5-Hexatriene	-0.099162	-0.081043	0	-0.209238	9.95	9.84
64	cis-1-3-5-Hexatriene	-0.096628	-0.072709	0	-0.209994	9.96	9.84
65	cis-Ocimene	-0.084047	-0.060893	0	-0.214641	9.98	9.85
66	2-Carene	-0.070809	-0.050403	0	-0.214811	10.00	9.81
67	cis-1-3-Pentadiene	-0.119947	-0.113836	0	-0.218090	10.00	10.01
68	1-3-5-Cycloheptatriene	-0.066484	-0.113412	0	-0.212641	10.01	9.81
69	2-Methyl-1-5-hexadiene	-0.079922	-0.111475	0	-0.241347	10.02	10.14
70	Cis-3-Methyl-2-pentene	-0.130291	-0.052227	0	-0.224349	10.03	10.07
71	1-Methyl-1-cyclopentene	-0.127674	-0.053838	0	-0.223283	10.04	10.05
72	2-Ethylbutene	-0.176768	-0.040028	0	-0.238794	10.05	10.34
73	D3-carene	-0.088817	-0.041613	0	-0.224523	10.06	9.96
74	trans-3-Hexene	-0.107116	-0.235035	0	-0.229750	10.08	10.15
75	cis-5-Decene	-0.114145	-0.242064	0	-0.231210	10.12	10.19
76	cis-3-Hexene	-0.112071	-0.243793	0	-0.231830	10.13	10.19
77	Methylketene	-0.189643	-0.179891	0	-0.222801	10.16	10.28
78	1,3-Butadiene	-0.132465	-0.134366	0	-0.228950	10.17	10.16
79	Cyclopentene	-0.111133	-0.249231	0	-0.232844	10.17	10.20
80	trans-2-Heptene	-0.109973	-0.237771	0	-0.230412	10.17	10.17
81	2-Methyl-1-pentene	-0.179815	-0.038736	0	-0.238959	10.20	10.35
82	1,5-Hexadiene	-0.081874	-0.123862	0	-0.249422	10.21	10.24
83	trans-2-Butene	-0.106220	-0.245143	0	-0.235027	10.22	10.21
84	3-Methyl-1,2-butadiene	-0.061527	-0.142706	0	-0.235227	10.24	10.05
85	2-Methyl-1-propene	-0.178806	-0.033716	0	-0.239567	10.26	10.35
86	cis-2-Butene	-0.113766	-0.256586	0	-0.233323	10.26	10.22
87	Camphene	-0.163119	-0.032538	0	-0.233314	10.27	10.24
88	2,3-Dimethyl-1-butene	-0.176179	-0.038128	0	-0.239239	10.28	10.34
89	Bicycle(2, 2, 1)-2-heptene	-0.117920	-0.251578	0	-0.230925	10.31	10.20
90	cis-Cyclooctene	-0.106120	-0.224366	0	-0.231830	10.38	10.17
91	Bicycle(2, 2, 2)-2-octene	-0.121856	-0.255257	0	-0.234580	10.39	10.25
92	4-Methyl-1-pentene	-0.145724	-0.201448	0	-0.249101	10.42	10.44
93	1-Decene	-0.106722	-0.164607	0	-0.246066	10.43	10.29
94	1-Butene	-0.157549	-0.236456	0	-0.246661	10.50	10.47
95	3,3-Dimethyl-1-butene	-0.156862	-0.211956	0	-0.249531	10.55	10.48
96	cis-1,3-Dichloropropene	-0.056614	-0.188993	1	-0.266567	11.08	11.13
97	1-Chloroethene	-0.142002	-0.171226	1	-0.262491	11.18	11.31
98	cis-1,2-Difluoroethene	-0.174089	-0.334662	2	-0.254139	12.15	12.14

Table 2. Molecular descriptors and $-\log k_{\text{OH}}$ values for 80 aromatics

No.	Name	E_{HOMO}	Q_{H}	$Mor17v$	Exp.	Pred.
Training set						
1	Toluene	-0.235310	0.161810	-0.152	11.22	11.33
2	N,N'-Dimethylaniline	-0.184365	0.162648	-0.090	9.83	9.71
3	m-Cresol	-0.215015	0.405793	-0.087	10.19	10.36
4	2,5-Dimethylphenol	-0.210349	0.406046	-0.152	10.10	10.08
5	2,3-Dichlorophenol	-0.237327	0.415574	0.109	11.78	11.49
6	m-Xylene	-0.228412	0.161007	-0.243	10.63	10.92
7	Bromobenzene	-0.241900	0.154319	-0.089	12.11	11.69
8	1,2,4-Trichlorobenzene	-0.254439	0.184731	0.102	12.26	12.44
9	p-Chloroaniline	-0.204776	0.322197	0.038	10.37	10.39
10	Benzonitrile	-0.266821	0.160515	-0.103	12.48	12.51
11	1,2,3-Trimethylbenzene	-0.226524	0.159394	-0.277	10.49	10.80
12	Biphenyl	-0.222169	0.132068	-0.172	11.14	10.90
13	2,2'-Dichlorobiphenyl	-0.243109	0.156510	-0.090	11.70	11.73
14	2,4-Dichlorobiphenyl	-0.236357	0.178026	-0.031	11.59	11.58
15	2',3,4-Trichlorobiphenyl	-0.241204	0.164889	0.152	12.00	12.12
16	2-Methylnaphthalene	-0.209415	0.161567	-0.182	10.28	10.39
17	1,4-Dichloronaphthalene	-0.224481	0.162327	-0.049	11.24	11.16
18	1,2,4-Trimethylbenzene	-0.221654	0.162312	-0.317	10.49	10.55
19	t-Butylbenzene	-0.235255	0.147250	-0.292	11.34	11.08
20	p-Cresol	-0.211012	0.405554	-0.136	10.33	10.13
21	2,4-Dichlorophenol	-0.233391	0.423640	0.063	11.97	11.25
22	1,2-Dihydroxy-4-methylbenzene	-0.202513	0.426353	-0.055	9.81	9.96
23	4-Chlorobenzotrifluoride	-0.264730	0.165715	-0.050	12.62	12.53
24	1,2-Dihydroxybenzene	-0.203874	0.406360	0.002	9.98	10.15
25	3,3'-Dichlorobiphenyl	-0.239759	0.157330	0.037	11.39	11.86
26	2,2',3,5'-Tetrachlorobiphenyl	-0.249698	0.169308	-0.001	12.10	12.11
27	2,3-Dimethylnaphthalene	-0.206205	0.159029	-0.271	10.11	10.11
28	p-Xylene	-0.225706	0.160255	-0.208	10.84	10.90
29	m-Ethyltoluene	-0.229391	0.160315	-0.369	10.72	10.71
30	1-Phenyl-2-methyl-1-propene	-0.212499	0.167845	-0.244	10.48	10.37
31	2,4-Dimethylphenol	-0.206656	0.412267	-0.209	10.15	9.83
32	p-Ethyltoluene	-0.226155	0.163692	-0.374	10.92	10.59
33	p-Dichlorobenzene	-0.247840	0.165042	0.037	12.49	12.12
34	Dibenzo-p-dioxin	-0.196206	0.147746	0.176	10.83	10.66
35	3,5-Dichlorobiphenyl	-0.240104	0.176852	0.001	11.38	11.77
36	p-Cymene	-0.226187	0.162981	-0.293	10.82	10.75
37	1,2-Dihydroxybenzene	-0.203874	0.406366	0.002	9.98	10.15
38	Benzyl alcohol	-0.241703	0.388807	-0.152	10.64	11.18
39	2-Naphthol	-0.205244	0.408348	-0.042	9.77	10.11
40	Acenaphthylene	-0.213398	0.134531	-0.462	9.96	10.03
Test set						
41	trans-1-Phenyl-1-propene	-0.212697	0.157210	-0.224	10.23	10.43
42	Phenol	-0.218943	0.406350	-0.043	10.58	10.58
43	2,3-Dimethylphenol	-0.211353	0.406310	-0.131	10.10	10.15

44	3,4-Dimethylphenol	-0.207982	0.159271	-0.164	10.09	10.38
45	1,2-Dihydroxy-3-methylbenzene	-0.200703	0.405596	-0.045	9.69	9.96
46	Fluorobenzene	-0.243342	0.146155	-0.042	12.16	11.85
47	m-Dichlorobenzene	-0.254309	0.178369	0.011	12.14	12.27
48	Aniline	-0.198155	0.317899	-0.029	9.95	10.05
49	Diphenylamine	-0.186785	0.144370	-0.038	9.71	9.93
50	Indane	-0.227288	0.150423	-0.289	11.04	10.81
51	2,3-Dihydrobenzofuran	-0.207605	0.164914	-0.224	10.44	10.24
52	3-Chlorobiphenyl	-0.231318	0.156809	-0.072	11.28	11.36
53	4,4'-Dichlorobiphenyl	-0.230332	0.157946	-0.071	11.70	11.33
54	2,4,5-Trichlorobiphenyl	-0.241571	0.184064	0.083	11.89	11.97
55	2,2',4,4'-Tetrachlorobiphenyl	-0.250343	0.180982	-0.024	12.00	12.06
56	1,3,5-Trimethylbenzene	-0.225937	0.160058	-0.314	10.24	10.70
57	o-Xylene	-0.229483	0.158403	-0.215	10.86	11.02
58	Isopropylbenzene	-0.235675	0.149636	-0.221	11.19	11.23
59	Methoxybenzene	-0.215159	0.168629	-0.030	10.76	10.87
60	2,6-Dimethylphenol	-0.210012	0.170326	-0.155	10.18	10.45
61	n-Propylbenzene	-0.235056	0.144776	-0.351	11.22	10.97
62	o-Dichlorobenzene	-0.251513	0.162814	0.066	12.38	12.31
63	2,4-Toluenediamine	-0.180661	0.320196	-0.034	9.72	9.43
64	2-Chlorobiphenyl	-0.232209	0.155215	-0.054	11.55	11.43
65	2,4,4'-Trichlorobiphenyl	-0.239363	0.180098	0.051	11.96	11.84
66	o-Ethyltoluene	-0.229576	0.161476	-0.362	10.91	10.73
67	Benzene	-0.246246	0.128582	-0.086	11.91	11.89
68	Ethylbenzene	-0.234059	0.153817	-0.230	11.15	11.15
69	4-t-Butyltoluene	-0.226003	0.160343	-0.339	10.86	10.66
70	o-Cresol	-0.213955	0.166169	-0.101	10.38	10.70
71	3,5-Dimethylphenol	-0.212285	0.405307	-0.158	9.95	10.13
72	Chlorobenzene	-0.246366	0.155025	-0.026	12.11	11.97
73	Tetralin	-0.228214	0.149512	-0.359	10.46	10.71
74	4-Chlorobiphenyl	-0.226420	0.155500	-0.124	11.41	11.10
75	2,4',5-Trichlorobiphenyl	-0.241429	0.166855	0.111	11.92	12.05
76	1-Methylnaphthalene	-0.208493	0.162456	-0.178	10.28	10.37
77	2-Phenyl-1-propene	-0.221340	0.161206	-0.222	10.28	10.72
78	1-Naphthol	-0.199275	0.409567	-0.050	9.26	9.89
79	2,3-Benzofuran	-0.219410	0.159895	-0.197	10.43	10.71
80	Styrene	-0.221666	0.144344	-0.130	10.24	10.94

2.2. Molecular descriptors

The molecular structure of each compound was sketched first using ChemBioDraw Ultra 11.0. Subsequently, the sketched structure was transferred to Chem3D module and pre-optimized using MM2 force field in ChemBio3D Ultra 11.0 until the minimum *rms* error became smaller than 0.100 kcal/mol Å. The energy minimized molecules were then fully optimized and calculated using the B3LYP approach in combination with the 6-31G(d) basis set,¹⁹ within Gaussian 09 (Revision A.02). Lastly, two groups of descriptors were calculated for each molecule. One comprises 1664 molecular descriptors that are based on the fully optimized molecular structures and calculated by Dragon software.²⁰ More information about the types of the molecular descriptors calculated with Dragon software can be found in Dragon software user's guide.²⁰ The other is the quantum chemical descriptors including the molecular average

polarizability (α), the molecular dipole moment (μ), the energy of the lowest unoccupied molecular orbital (E_{LUMO}), the energy of the highest occupied molecular orbital (E_{HOMO}), the most positive net atomic charge on hydrogen atoms in a molecule (Q_{H}), the net charge of the most negative atom (q) and Fukui indices. The former six descriptors (α , μ , E_{LUMO} , E_{HOMO} , Q_{H} , and q) are presented in the modified theoretical linear solvation energy relationship model.²¹ For 98 alkenes $\text{C}_1\text{R}_3\text{R}_4=\text{C}_2\text{R}_5\text{R}_6$ (C_1 contains less pendent groups than C_2), three types of Fukui indices were calculated for C_1 and C_2 atoms with following equations:^{22,23}

$$f_r^- = q_r^N - q_r^{N-1} \quad (\text{for electrophilic attack}) \quad (1)$$

$$f_r^+ = q_r^{N+1} - q_r^N \quad (\text{for nucleophilic attack}) \quad (2)$$

$$f_r^0 = (q_r^{N+1} - q_r^{N-1}) / 2 \quad (\text{for radical attack}) \quad (3)$$

where r is the atom C_1 or C_2 , q_r^N , q_r^{N-1} and q_r^{N+1} are, respectively, Mulliken atom charges of the neutral, cationic, and anionic species. Thus six Fukui indices were calculated for each alkene. Similarly, 21 Fukui indices for seven atoms were calculated for each aromatic compound. The seven atoms include the six C atoms in the benzene ring and the atom in substituents, which joins directly to the ring.

2.3. Model development

MLR was used to seek an optimum linear combination of variables from the descriptors calculated and develop a MLR model.^{24,25} Some important statistical parameters, such as the correlation coefficient R , standard error SE , t -test, variance inflation factor (VIF) and $Sig.$ -test (or p -value), were used to evaluate the variables or models.

The leave-one-out (LOO) cross-validation procedure can be used to test the internally predictive ability of a QSAR model. In LOO each chemical is put in the test set once a time in every iterative validation: thus it is a test compound when it is predicted by the model developed on $n-1$ chemicals. A successful QSPR model should be validated with the test set and satisfies following criteria^{26,27}

$$q_{\text{int}}^2 > 0.5 \quad (4)$$

$$q_{\text{ext}}^2 > 0.5 \quad (5)$$

$$R^2 > 0.6 \quad (6)$$

$$(R^2 - R_0^2) / R^2 < 0.1 \text{ or } (R^2 - R_0^2) / R^2 < 0.1 \quad (7)$$

$$0.85 \leq k \leq 1.15 \text{ or } 0.85 \leq k' \leq 1.15 \quad (8)$$

where q^2 is the predictive squared correlation coefficient (q_{int}^2 for training sets, q_{ext}^2 for external test sets), R is the correlation coefficient, R_0^2 (predicted versus observed values) and $R_0'^2$ (observed versus predicted values) are coefficients of determination, k and k' are slopes of regression lines through the origin of predicted versus observed and observed versus predicted respectively. Mathematical definitions of parameters (q^2 , R^2 , R_0^2 , $R_0'^2$, k and k') can be found in the literature.^{27,28}

3. RESULTS AND DISCUSSION

3.1. Statistical results

An analysis of the rate constants $-\log k_{\text{OH}}$ in the training sets of alkenes and aromatics with respect to respective descriptors resulted in the following best regression equations:

$$-\log k_{\text{OH}} = 7.346 - 2.671 f_{1C}^- - 0.486 f_{2C}^- + 0.749 N_X - 10.481 E_{\text{HOMO}} \quad (9)$$

$R^2 = 0.938$, $s = 0.108$, $F = 166.530$, $N = 49$ (For alkenes)

$$-\log k_{\text{OH}} = 3.865 - 34.141 E_{\text{HOMO}} - 1.660 Q_H + 1.943 \text{Mor17v} \quad (10)$$

$R^2 = 0.910$, $s = 0.265$, $F = 121.212$, $N = 40$ (For aromatics)

The descriptors present in above models and corresponding values are shown in Tables 1 and 2; their meanings and statistical parameters are shown in Table 3. The results evaluated are listed in Table 4. Calculated values are listed in Tables 1 and 2, and plotted versus experimental values in Figs 1 and 2.

The $Sig.$ -values in Table 3 suggests that all the descriptors in each model are significant; the VIF -test shows that the descriptors are not strongly correlated with each other. From Table 4, we can conclude that our results satisfy the above accept conditions (Eqs 4–8).

3.2. Comparison of models

In this paper, the rms errors of the training and test sets of alkenes are 0.102 and 0.114, respectively. The mean rms error is 0.108, which is closer to the results (0.106 ~ 0.139) in other MLRs of alkenes.^{7,8} But our model includes more compounds and fewer descriptors. In addition, our model (Eq. 9) and Fatemi's models¹² based on alkenes have the same data set. But the rms errors of the six QSAR models, based on five descriptors and evaluated by a leave-24-out cross-validation test, were above 0.16.¹² Compared with other models in the literature,^{7,8,12} the present MLR model (Eq. 9) shows better

statistical quality.

The rms errors of the training and test sets for aromatics in this paper are 0.282 and 0.260, respectively. While the rms errors in previous models^{9-11,13} containing aromatics and non-aromatics range between 0.391 and 0.501, which are greater than the results in our model (Eq. 10). In addition, the squared correlation coefficient ($R^2 = 0.910$) in Eq. 10 is greater than that ($r^2 = 0.735$) of the 4-variable model of k_{OH} for aromatic pollutants in water matrix,²⁹ which also suggests that our model (Eq. 10) is accurate.

Generally, a QSAR model of k_{OH} , with rms values of around 0.35 log units, can be considered acceptable for screening purposes.³⁰ Obviously, the models obtained in this paper satisfy this criterion.

Table 3. List of descriptors in MLR models, meaning and statistical parameters.

Model	Symbol	Descriptor	t -test	$Sig.$	VIF
Model for alkenes	f_{1C}^-	The Fukui index of C_1 atom based on Mulliken charges	-5.871	0.000	1.459
	f_{2C}^-	The Fukui index of C_2 atom based on Mulliken charges with hydrogens summed into heavy atoms	-2.454	0.018	1.331
	N_X	The number of R=CHX functional group counts	13.928	0.000	1.192
	E_{HOMO}	The energy of the highest occupied molecular orbital	-6.811	0.000	1.913
Model for aromatics	E_{HOMO}	The energy of the highest occupied molecular orbital	-13.693	0.000	1.238
	Q_H	The most positive net atomic charge on hydrogen atoms in a molecule	-3.879	0.000	1.313
	Mor17v	3D-MorSE descriptor (signal 17 / weighted by atomic van der Waals volumes)	6.429	0.000	1.180

3.3 Applicability domain

The Williams plots (i.e. plot of standardized residuals versus leverages) in Figs 3 and 4 were used to visualize respective applicability domains. Predictions for only those compounds that fall into this domain may be considered reliable.^{13,25} The leverage h and warning leverage h^* are defined with the following expressions^{13,31}

$$h_i = x_i^T (X^T X)^{-1} x_i \quad (i = 1, \dots, n) \quad (11)$$

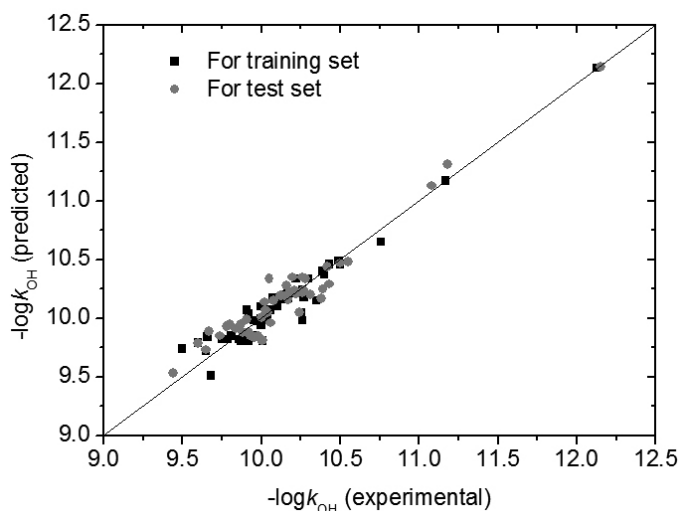
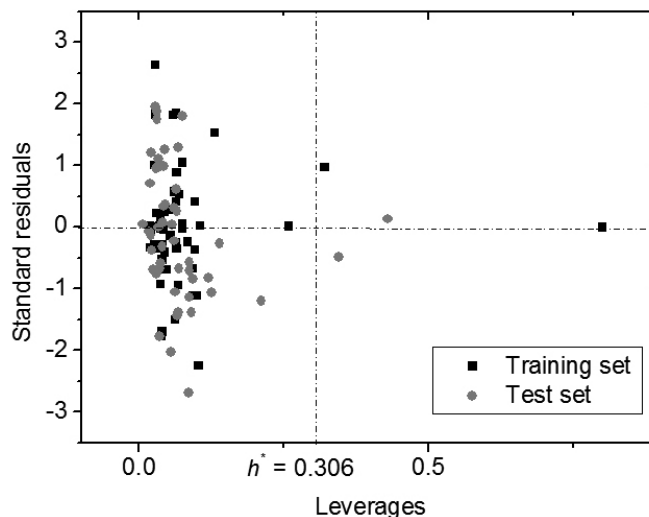
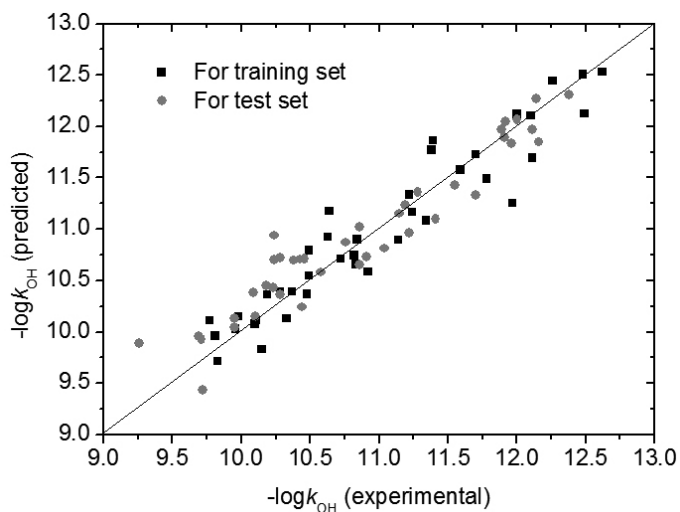
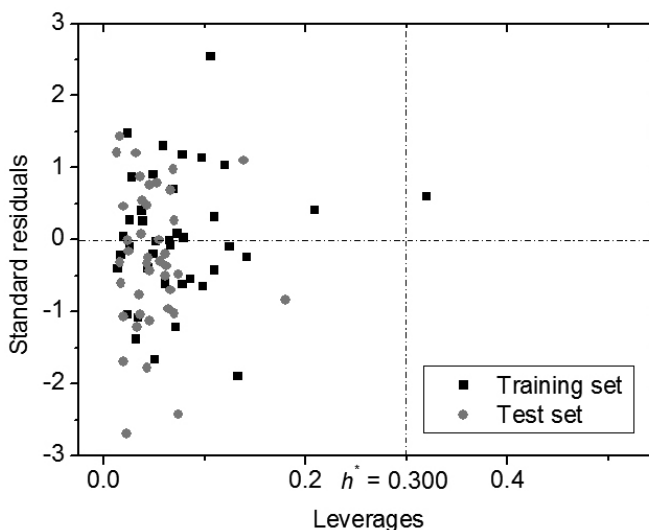
$$h^* = 3(p + 1) / n \quad (12)$$

where x_i is the descriptor vector of the considered compound, X is the model matrix derived from the training set descriptor values, n is the number of training compounds and p is the number of model parameters.

Generally, a value of 3 for standardized residual is used as a cut-off value for accepting predictions. High leverage points ($h_i > h^*$) with small standardized residuals ($< 3\sigma$) are taken as good high leverage points or good influence points, which stabilize the model and make it more precise. While high leverage points with large standardized residuals ($> 3\sigma$) are called bad high leverage points or bad influence points.³¹ For alkenes, Fig. 3 shows that two chemicals in the training set have leverage values greater than the warning leverage h^* ($= 0.306$). But their σ values are less than 3. Thus these points can stabilize the model and make it more accurate. In addition, the two chemicals with $h > h^*$ and $\sigma < 3$ in the test set suggest that the MLR model of alkenes has good generalizability.^{13,31} Similarly, as can be seen from Fig. 4, there is only one chemical with $h > h^*$ ($= 0.300$) and $\sigma < 3$ in the training set, which exerts a large influence on the model of aromatics.

Table 4. Results evaluated for the two models

Model	q_{int}^2	q_{ext}^2	R^2	R_0^2	$R_0'^2$	k	k'	rms
Model for alkenes	0.925	0.928	0.930	0.927	0.929	1.001	0.999	0.108
Model for aromatics	0.886	0.903	0.916	0.879	0.908	1.005	0.994	0.260

**Figure 1.** Plot of calculated vs. experimental $-\log k_{\text{OH}}$ values of alkenes.**Figure 3.** Williams plot for the model of alkenes.**Figure 2.** Plot of calculated vs. experimental $-\log k_{\text{OH}}$ values of aromatics.**Figure 4.** Williams plot for the model of aromatics.

3.4. Mechanistic interpretation

For the degradation reactions of alkenes, OH radical addition to double bond is the dominant reaction pathway: $\cdot\text{OH} + \text{C}_1\text{R}_3\text{R}_4=\text{C}_2\text{R}_5\text{R}_6 \rightarrow \text{C}_1(\text{OH})\text{R}_3\text{R}_4-\text{C}_2\text{R}_5\text{R}_6$. According to the classic structure-activity relationship, the chemical reactivity of vinyl compounds is determined by electronic and steric effects of substituents. Electronic effects include inductive and conjugated effects.²⁹ Steric effects include the size, position and number of substituents. Electron-donor substituents of alkenes can increase the electron density of double bonds, activate π bonds, and result in high degradation rates for electrophilic attack by OH radical. In addition, conjugated double bonds of alkenes increase electron mobility, inductive effect, molecular polarizability and degradation rates. Therefore, from Table 1, we can find the trend that dialkenes have higher rate constants than monoolefines. The synergistic effects of electron-donor substituents and conjugated double bonds increase reaction rates. For example, α -Phellandrene ($-\log k_{\text{OH}} = 9.50$) (No. 1 in Table 1) and α -Terpinene

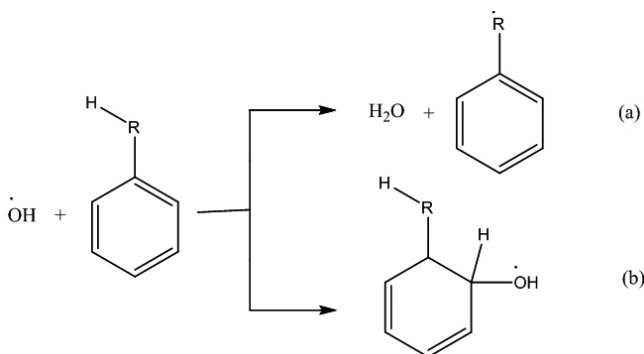
($-\log k_{\text{OH}} = 9.44$) (No. 50) are the most reactive systems and have the highest rate constants. On the contrary, electron-acceptor groups of alkenes such as halogen and carbonyl substituents deactivate double bonds and consequently decrease the degradation of chemicals. For example, 1-Bromoethene (No. 48 in Table 1), trans-1,2-Difluoroethene (No. 49), 1-Chloroethene (No. 97), and cis-1,2-Difluoroethene (No. 98) are the least reactive molecules and have the lowest rate constant values.

By the t -test (see Table 3), the most significant descriptor in Eq. 9 is N_{X} , which denotes the number of $\text{R}=\text{CHX}$ functional group and reflects the number of halogen atoms.²⁰ Halohydrocarbons are classic electrophilic reagents, which are not susceptible to electrophilic attack. Therefore, the molecules with more $\text{R}=\text{CHX}$ groups tend to have lower k_{OH} values. The second significant descriptor appearing in Eq. 9 is E_{HOMO} . The descriptor E_{HOMO} governs many chemical reactions and characterizes the susceptibility of the molecule toward attack by electrophiles.⁶ The OH radical is an electrophilic reagent during the

reactions of alkenes with $\cdot\text{OH}$ by the radical addition to the $>\text{C}=\text{C}<$ double bond. An alkene compound with a larger E_{HOMO} tends to lose electrons and leads to increased susceptibility of $\cdot\text{OH}$ attacking and consequently a high k_{OH} value.^{13,15} Thus, the descriptor E_{HOMO} is positively correlated with k_{OH} .

According to the *t*-test, the next two significant descriptors $f_{1\text{C}}$ and $f'_{2\text{C}}$ are Fukui indices of C_1 and C_2 atoms for electrophilic attack, respectively. Fukui indices are very useful in characterizing the location of the reactive regions as well as either the nucleophilic or electrophilic behavior. A larger value of Fukui index $f_{1\text{C}}$ (or $f'_{2\text{C}}$) indicates a higher reactivity for C_1 (or C_2) atom. Fukui indices $f_{1\text{C}}$ and $f'_{2\text{C}}$ are correlated with k_{OH} , which suggests that both C_1 and C_2 exhibit high reactivity when OH radical attacks double bonds of alkenes.

As a strong electrophilic species, OH radical reacts with aromatic compounds by following two routs (Scheme 1).¹ One is H-atom abstraction from the C–H bond of alkyl substituent group (Scheme 1 (a)). The other is the OH radical addition to one of double bonds from the benzene rings (Scheme 1 (b)). Above all, we take aromatic compounds with only one substituent as examples to analyze substituent effects. Aromatics with electron-donor groups $-\text{N}(\text{CH}_3)_2$ (9.83), $-\text{NH}_2$ (9.95), $-\text{OH}$ (10.58), $-\text{OCH}_3$ (10.76), $-\text{CH}_2\text{CH}_3$ (11.15), $-\text{CH}(\text{CH}_3)_2$ (11.19), $-\text{CH}_3$ (11.22), $-\text{CH}_2\text{CH}_2\text{CH}_3$ (11.22), $-\text{C}(\text{CH}_3)_3$ (11.34) possess higher degradation rate k_{OH} values (or lower $-\log k_{\text{OH}}$ values) than Benzene (11.91). The above values in the parentheses are the corresponding $-\log k_{\text{OH}}$ values. While the compound Benzonitrile with an electron-acceptor substituent $-\text{CN}$ has a low k_{OH} value (12.48). The same trends can be found in other aromatic compounds. These phenomena can be explained by electronic effects of substituents, i.e., electron-donor substituents can activate the benzene ring and increase susceptibility of $\cdot\text{OH}$ attacking, and on the contrary, electron-acceptor substituents shift the electron density and deactivate the benzene ring.²⁹ In addition, Naphthalenes have greater molecular polarizability and higher reaction activity than Benzene. Therefore, it is easy to understand that the compound 1-Naphthol with the strong electron-donor group $-\text{OH}$ shows the highest k_{OH} value ($-\log k_{\text{OH}} = 9.26$) among 80 aromatics (see Table 2). All the halogen derivatives have low k_{OH} values, especially for these compounds with multiple halogen substitutions. For example, 4-Chlorobenzotrifluoride ($-\log k_{\text{OH}} = 12.62$) and *p*-Dichlorobenzene ($-\log k_{\text{OH}} = 12.49$) show the lowest k_{OH} values. The reasons are that halogen substituents are electron-acceptor groups and multiple halogen groups have synergistic effects on deactivating benzene rings, and furthermore, decreasing k_{OH} values.



Scheme 1. The reaction routs for aromatic compounds with OH radical.

For the model (Eq. 10) of aromatics, the most significant descriptor is E_{HOMO} . As stated above, E_{HOMO} is positively related to k_{OH} . The next one is the 3D MoRSE (3D Molecule Representation of Structures based on Electron diffraction) descriptor *Mor17v* (signal 17/weighted by atomic van der Waals volumes). *Mor17v* retains important structural features such as the mass and the amount of branching.²⁰ An increase in *Mor17v* causes an increase in the steric hindrance between the molecule and OH radical. Therefore, the degradation rate constant decreases. The last significant descriptor in Eq. 10 is Q_{H} . It denotes the most positive net atomic charge on hydrogen atoms. Generally, local electron densities or charges are used as static chemical reactivity indices and describe molecular polarizability.⁶ The greater the net charge on H atom is, the higher the reactivity. Thus k_{OH} increases with increasing Q_{H} .

4. CONCLUSIONS

General QSAR models based on the MLR approach were developed for reaction rate constants of $\cdot\text{OH}$ with 98 alkenes and 80 aromatics in the atmosphere. The two QSAR models were proved to be accurate and reliable.

For the model for k_{OH} of alkenes, the most significant descriptor is the number of $\text{R}=\text{CHX}$ functional group counts, N_{X} . The group deactivates the double bond and decreases the reaction rate. The second descriptor E_{HOMO} characterizes the susceptibility of the molecule toward attack by electrophiles. A larger E_{HOMO} tends to lose electrons and facilitates the degradation of chemicals. Fukui indices of C_1 and C_2 atoms ($f_{1\text{C}}$ and $f'_{2\text{C}}$) characterize the location of the reactive regions for electrophilic attack. A larger value of Fukui index $f_{1\text{C}}$ (or $f'_{2\text{C}}$) indicates a higher reactivity for C_1 (or C_2) atom.

For the model for k_{OH} of aromatics, the main contribution to degradation rates is E_{HOMO} . A larger E_{HOMO} results in a higher k_{OH} . The topological descriptor *Mor17v* encodes molecular structure and steric hindrance. A larger descriptor *Mor17v* indicates greater steric hindrance and consequently decreases the reaction rate constant. Q_{H} can reflect molecular reactivity and polarizability. Thus, the descriptors in each MLR model have clear physical meanings and represent respective major factors affecting the degradation rates k_{OH} .

ACKNOWLEDGEMENTS

We thank the financial supports from the Science Foundation of Hunan Province (Grant No. 2010FJ4116), Hunan Provincial Natural Science Foundation of China (Grant No. 12JJ6011), and the National Natural Science Foundation of China (Grant No. 20972045).

REFERENCES

1. R. Atkinson, *Atmos. Environ.* **41**, S200, (2007).
2. P. Gramatica, V. Consonni, R. Todeschini, *Chemosphere* **38**, 1371, (1999).
3. P. Gramatica, P. Pilutti, E. Papa, *SAR QSAR Environ. Res.* **13**, 743, (2002).
4. W. M. Meylan, P. H. Howard, *Environ. Toxicol. Chem.* **22**, 1724, (2003).
5. A. Böhnhardt, R. Kühne, R.-U. Ebert, G. Schüürmann, *Theor. Chem. Acc.* **127**, 355, (2010).
6. M. Karelson, V. S. Lobanov, A. R. Katritzky, *Chem. Rev.* **96**, 1027, (1996).
7. G. A. Bakken, P. C. Jurs, *J. Chem. Inf. Comput. Sci.* **39**, 1064, (1999).
8. M. Pompe, M. Veber, M. Randić, A. T. Balaban, *Molecules* **9**, 1160, (2004).
9. P. Gramatica, P. Pilutti, E. Papa, *Atmos. Environ.* **38**, 6167, (2004).
10. P. Gramatica, P. Pilutti, E. Papa, *J. Chem. Inf. Comput. Sci.* **44**, 1794, (2004).
11. T. Öberg, *Atmos. Environ.* **39**, 2189, (2005).
12. M. H. Fatemi, E. Baher, *SAR QSAR Environ. Res.* **20**, 77, (2009).
13. Y. N. Wang, J. W. Chen, X. H. Li, B. Wang, X. Y. Cai, L. P. Huang, *Atmos. Environ.* **43**, 1131, (2009).
14. C. L. Yan, J. W. Chen, L. P. Huang, G. H. Ding, X. Y. Huang, *Chemosphere* **61**, 1523, (2005).
15. X. X. Long, J. F. Niu, *Chemosphere* **67**, 2028, (2007).
16. P. Gramatica, P. Pilutti, E. Papa, *Atmos. Environ.* **37**, 3115, (2003).
17. M. Pompe, M. Veber, *Atmos. Environ.* **35**, 3781, (2001).
18. M. H. Fatemi, *Anal. Chim. Acta* **556**, 355, (2006).
19. X. L. Yu, B. Yi, X. Y. Wang, Z. M. Xie, *Chem. Phys.* **332**, 115, (2007).
20. R. Todeschini, V. Consonni, A. Mauri, M. Pavan, *DRAGON Professional*, Version 5.4. Talete srl: Milan, 2006.
21. J. W. Chen, L. Feng, Y. Y. Liao, S. K. Han, L. S. Wang, *Chemosphere* **33**, 537, (1996).
22. R. G. Parr, W. Yang, *J. Am. Chem. Soc.* **106**, 4049, (1984).
23. B. Gómez, N. V. Likhanova, M. A. Domínguez Aguilar, O. Olivares, J. M. Hallen, J. M. Martínez-Magádn, *J. Phys. Chem. A* **109**, 8950, (2005).
24. X. L. Yu, B. Yi, X. Y. Wang, *J. Comput. Chem.* **28**, 2336, (2007).
25. X. L. Yu, B. Yi, X. Y. Wang, J. F. Chen, *Atmos. Environ.* **51**, 124, (2012).
26. A. Golbraikh, A. Tropsha, *J. Mol. Graph. Model.* **20**, 269, (2002).
27. A. Tropsha, P. Gramatica, V. K. Gombar, *QSAR Combi. Sci.* **22**, 69, (2003).
28. X. L. Yu, B. Yi, W. H. Yu, X. Y. Wang, *Chem. Pap.* **62**, 623, (2008).
29. H. Kušić, B. Rasulev, D. Leszczynska, J. Leszczynski, N. Koprivanac, *Chemosphere* **75**, 1128, (2009).
30. A. Böhnhardt, R. Kühne, R.-U. Ebert, G. Schüürmann, *J. Phys. Chem. A* **112**, 11391, (2008).
31. L. Eriksson, J. Jaworska, A. P. Worth, M. T. D. Cronin, R. M. McDowell, P. Gramatica, *Environ. Health Perspect.* **111**, 1361, (2003).