Available online at www.sciencedirect.com
ScienceDirect
Toxicology Letters 175 (2007) 136–144
Toxicology Letters
www.elsevier.com/locate/toxlet

Prediction of retention times for a large set of pesticides or toxicants based on support vector machine and the heuristic method

Xiuyong Li [a,b], Feng Luan [a], Hongzong Si [c], Zhide Hu [a,*], Mancang Liu [a]

[a] *Department of Chemistry, Lanzhou University, Lanzhou, Gansu 730000, China*
[b] *Yantai Entry-Exit Inspection and Quarantine Administration, Yantai, Shandong 264000, China*
[c] *Institute for Computational Science and Engineering, Qingdao University, Qingdao, Shandong 266071, China*

Received 2 July 2007; received in revised form 11 October 2007; accepted 11 October 2007
Available online 18 October 2007

Abstract

Quantitative structure–retention relationship (QSRR) studies were performed for predicting the retention times (RTs) of 110 kinds of pesticides or toxicants. Chemical descriptors were calculated from the molecular structure of the compounds alone. The QSRR models were built using the heuristic method (HM) and support vector machine (SVM), respectively. The obtained linear model of HM had a square of a correlation coefficient: $R^2 = 0.913$, $F = 116.70$ with a root mean square error (RMS) error of 0.0387 for the training set, while $R^2 = 0.907$, $F = 195.49$, and RMS = 0.0408 for the test set. The non-linear model by SVM gave better results: for the training set $R^2 = 0.966$, $F = 2420.5$, RMS = 0.0231 and for the test set $R^2 = 0.944$, $F = 339.7$, RMS = 0.0313. The prediction results are in good agreement with the experimental values. And the proposed model could identify and provide some insight into what structural features are related to retention time of these compounds.

Keywords: Pesticides or toxicants; QSRR; Retention times; Molecular descriptors; SVM

1. Introduction

The study of quantitative structure–retention relationship (QSRR) of solutes is an important topic in chromatographic thermodynamics. It belongs to the most often studied manifestations of the linear free-energy relationships (LEFR) (Kaliszan, 1987). QSRR are the statistically derived relationships between the chromatographic parameters determined for a structurally diverse series of analytes in a given separation system and the descriptors accounting for the structural differences among the analytes studied (Al-Haj et al., 1999). QSRR provides a promising method for the estimation of retention times (RTs) based on descriptors derived solely from the molecular structure to fit experimental data. The advantages of this approach lie in the fact that it requires the knowledge of chemical structure and is dependent on few experiment data.

Over the past several decades, QSRR studies are widely investigated. Correlation between chromatographic retention indices and molecular parameters provided significant information: on the effect of the molecular structure on retention time and on the possible mechanism of absorption and elution (KÖrtvélyesi et al., 2001). This could be achieved using quantitative structure with the retention phenomena (Olivero et al.,

* Corresponding author.
E-mail address: lxyyt65@sina.com (Z. Hu).

0378-4274/$ – see front matter © 2007 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.toxlet.2007.10.005

Table 1
Pesticides or toxicants, observed and calculated values of Retention Time

		Observed log t_R		Calculated log t_R		
		log t_R	HM	Residue	SVM	Residue
1[a]	Ethoprophos	1.195	1.247	0.052	1.2317	0.0367
2	Demeton-*s*-methyl	1.217	1.2325	0.0155	1.2168	−0.0002
3	Omethoate	1.239	1.2776	0.0386	1.2392	0.0002
4	Phorate	1.247	1.3061	0.0591	1.3127	0.0657
5	a-666	1.275	1.2893	0.0143	1.2748	−0.0002
6[a]	Terbufos	1.303	1.32	0.017	1.3386	0.0356
7	Chlorbufan	1.313	1.3125	−0.0005	1.3132	0.0002
8	Atrazine	1.323	1.3468	0.0238	1.3228	−0.0002
9	Trietazine	1.326	1.3606	0.0346	1.3262	0.0002
10	Fonofos	1.337	1.3646	0.0276	1.3474	0.0104
11[a]	Lindan	1.34	1.2893	−0.0507	1.2748	−0.0652
12	PCB15	1.344	1.4279	0.0839	1.3442	0.0002
13	Disulfoton	1.35	1.3319	−0.0181	1.3498	−0.0002
14	Dimethoate	1.356	1.3512	−0.0048	1.3562	0.0002
15	Carbofuran	1.36	1.3597	−0.0003	1.3598	−0.0002
16[a]	Dichlofenthion	1.387	1.4841	0.0971	1.3957	0.0087
17	4,4′-DDM	1.39	1.4724	0.0824	1.3902	0.0002
18	PCB31	1.399	1.4676	0.0686	1.4174	0.0184
19	Benoxacor	1.401	1.3925	−0.0085	1.4012	0.0002
20	Fenchlorphos	1.429	1.5056	0.0766	1.504	0.075
21[a]	Phosphamidon	1.43	1.3462	−0.0838	1.3839	−0.0461
22	Benfuresate	1.432	1.4291	−0.0029	1.4319	−0.0001
23	Aldrin	1.434	1.4978	0.0638	1.4927	0.0587
24	PCB52	1.443	1.5103	0.0673	1.4812	0.0382
25	Parathion-methyl	1.449	1.4201	−0.0289	1.3963	−0.0527
26[a]	Metalaxyl	1.449	1.4691	0.0201	1.4877	0.0387
27	Pentanochlor	1.451	1.3789	−0.0721	1.393	−0.058
28	Pirimiphos-methyl	1.452	1.4741	0.0221	1.4731	0.0211
29	Paraoxon-ethyl	1.463	1.4354	−0.0276	1.4511	−0.0119
30	Metolachlor	1.464	1.459	−0.005	1.48	0.016
31[a]	Trichoronate	1.472	1.4999	0.0279	1.5086	0.0367
32	Methoprene	1.476	1.5084	0.0324	1.4762	0.0002
33	Chlorpyriphos	1.476	1.5296	0.0536	1.5184	0.0424
34	Fenitrothion	1.477	1.483	0.006	1.4917	0.0147
35	Malathion	1.478	1.4887	0.0107	1.4779	−0.0002
36[a]	Thiobencarb	1.478	1.4433	−0.0347	1.4754	−0.0026
37	Methiocarb	1.483	1.384	−0.099	1.3971	−0.0859
38	Isodrin	1.485	1.4932	0.0082	1.4848	−0.0002
39	Parathion	1.495	1.4745	−0.0205	1.4948	−0.0002
40	Pirimiphod	1.495	1.474	−0.021	1.4766	−0.0184
41[a]	Fenthion	1.501	1.4888	−0.0122	1.5158	0.0148
42	Allethrin	1.503	1.5161	0.0131	1.5102	0.0072
43	Bromophos-methyl	1.504	1.513	0.009	1.5138	0.0098
44	Pendimethalin	1.516	1.5165	0.0005	1.5336	0.0176
45	Isocarbophos	1.52	1.4802	−0.0398	1.5105	−0.0095
46[a]	PCB70	1.522	1.5315	0.0095	1.5115	−0.0105
47	Isofenphos	1.523	1.5408	0.0178	1.5516	0.0286
48	Flumetralin	1.533	1.5666	0.0336	1.5332	0.0002
49	Triadimenol	1.533	1.5596	0.0266	1.5731	0.0401
50	Bromophos-ethyl	1.539	1.5259	−0.0131	1.5392	0.0002
51[a]	Chlorfenvinphos	1.54	1.5221	−0.0179	1.535	−0.005
52	Procymidone	1.544	1.5584	0.0144	1.5441	0.0001
53	PCB101	1.545	1.5622	0.0172	1.5575	0.0125
54	2,4′-DDE	1.545	1.6019	0.0569	1.6206	0.0756
55	Quinalophos	1.549	1.5413	−0.0077	1.5775	0.0285
56[a]	Alpha-endosulfan	1.549	1.5703	0.0213	1.5624	0.0134