Journal of Environmental Sciences Vol.16, No.1, pp.21—29, 2004

ISSN 1001-0742
CN11-2629/X

Article ID: 1001-0742(2004)01-0021-09 **CLC number**: X132 **Document code**: A

Application of TLSER method in predicting the aqueous solubility and *n*-octanol/water partition coefficient of PCBs, PCDDs and PCDFs

HUANG Jun[1], YU Gang[1,*], ZHANG Zu-lin[1], WANG Yi-lei[2], ZHU Wei-hua[2], WU Guo-shi[2]

(1. Department of Environmental Science and Engineering, POPs Research Center, Tsinghua University, Beijing 100084, China. E-mail: yg-den@tsinghua.edu.cn; 2. Department of Chemistry, Tsinghua University, Beijing 100084, China)

Abstract: The theoretical linear solvation energy relationship (TLSER) approach was adopted to predict the aqueous solubility and *n*-octanol/water partition coefficient of three groups of environmentally important chemicals-polychlorinated biphenyls (PCBs), polychlorinated dibenzodioxins and dibenzofurans (PCDDs and PCDFs). For each compound, five quantum parameters were calculated using AM1 semiempirical molecular orbital methods and used as structure descriptors: average molecular polarizability (α), energy of the lowest unoccupied molecular orbit (E_{LUMO}), energy of the highest occupied molecular orbit (E_{HOMO}), the most positive charge on a hydrogen atom (q_+), and the most negative atomic partial charge (q_-) in the solute molecule. Then standard independent variables in TLSER equation was extracted and two series of quantitative equations between these quantum parameters and aqueous solubility and *n*-octanol/water partition coefficient were obtained by stepwise multiple linear regression (MLR) method. The developed equations have both quite high accuracy and explicit meanings. And the cross-validation test illustrated the good predictive power and stability of the established models. The results showed that TLSER could be used as a promising approach in the estimation of partition and solubility properties of macromolecular chemicals, such as persistent organic pollutants.

Keywords: persistent organic pollutants (POPs); quantitative structure property relationship (QSPR); theoretical linear solvation energy relationship (TLSER); polychlorinated biphenyls (PCBs); polychlorinated dibenzo-*p*-dioxins (PCDDs); polychlorinated dibenzofurans (PCDFs)

Introduction

Polychlorinated biphenyls (PCBs), polychlorinated dibenzo-*p*-dioxins and dibenzofurans (PCDDs and PCDFs) play an important role as environmental contaminants, which have been included in many blacklists such as "persistent organic pollutants (POPs)" and suspected "environmental endocrine disruptors (EEDs)" (UNEP, 2001; Kavlock, 1996). The concentrations of these chemicals are quite low or even imperceptible, but still pose a serious health risk to humans and animal species, not only due to their high toxicity and other great diverse effects, but also due to their long persistence and long-range transport potentials.

For organic micropollutants, partitioning properties and aqueous solubility play a major role in their transport and mobility in the global environments, which is particularly true for POPs, such as PCBs, PCDDs and PCDFs. Consequently, it is very important to get these data accurately for the evaluation of humane exposure and risk assessment. However, all these three kinds of chemicals found in the environment are complex mixtures of large number congeners. According to the difference of number and position of chlorine atoms within the molecule, PCBs has 209 possible structure patterns, while 75 for PCDDs and 135 for PCDFs. This complexity had caused the difficulty both for the qualitative and for the quantitative analysis, and so does the determination of partitioning properties and aqueous solubility (Van Emon, 2000). Traditional experimental determination methods usually need special equipment and samples, cost large amounts of money, time and manpower. As a result, these values can be found only for a limited number of PCBs, PCDDs and PCDFs currently (Mackay, 1997), which could not meet the needs of the risk assessment of PCBs, PCDDs and PCDFs. Thus, the quantitative structure-property relationship (QSPR) method, which correlate and predict property data of pollutants from their structural descriptors, may be used to study the partition coefficients or aqueous solubility, and generate predicted data efficiently.

Up to now, a large number of calculation methods have been developed for the estimation of the partition coefficient and aqueous solubility with varying success and applicability. According to the molecular descriptors scientists used, these methods can be classified into two groups: empirical method and theoretical method. Some papers using empirical molecular descriptors, such as gas chromatography retention index (GC-RI) (Wang, 2002), SOFA (solubility parameters for fate analysis) model descriptors (Govers, 1995a; 1995b; 1998), mobile order thermodynamics (MOD) model descriptors (Rulle, 1997; 2000). However, the accurate determination of empirical molecular descriptors may be difficult, expensive in terms of cost and time, or even impossible for some PCBs, PCDDs and PCDFs, which might not have synthesized or purified. Also the experimental errors may be introduced, especially for those congeners difficult to be separated and identified by chromatography. Furthermore, the empirical molecular descriptors often reflect complex and multiple physical interactions. Hence, the interpretation of the respective QSPR could be difficult and ambiguous.

As an alternative, the theoretical descriptors derived solely from the chemical structure and the quantum mechanical calculated wave function of a molecule could, in principle, be obtained for any individual molecular structure. The theoretical descriptors have, in most cases, explicit mathematical definition based on fundamental physical equations of molecular matter or consistent physical models. Therefore, the QSPR equations based on theoretical descriptors embody distinct physical meaning related to clearly defined physical molecular models and model

Foundation item: The National Key Basic Research Foundation of China (No. G1999045711); * Corresponding author

processes. These features of the theoretical molecular descriptors make them particularly attractive and advantageous in the QSPR study. Among many theoretical molecular descriptors, quantum chemical descriptors are of great concerns in recent years as they can express, in principle, almost any electronic and geometric property of molecules and characteristics of intermolecular interactions and they are not restricted to closely related compounds(Karelson, 2000).

In this study, TLSER(theoretical linear solvation energy relationship) methodology developed by Famini and Wilson was applied to obtain the correlation equations (Famini, 1992; Wilson, 1991), based on five quantum chemical parameters calculated by semiempirical molecular orbital methods and molecular van de Waals volume parameter. Compared with recently published literature also using quantum descriptors(Chen, 2001), this study used an expanded sample set and the yield equations embodied more distinct physical meaning.

1 Materials and method

The sample sets organized by Rulle *et al*. (Rulle, 1997; 2000) were adopted in this study. For aqueous solubility ($-\log S_w$), there are 63 samples in PCBs set, 14 samples in PCDDs set and 7 samples in PCDFs set. For *n*-octanol/water partition coefficient($\log K_{ow}$), there are 154 samples in PCBs set, 42 samples in PCDDs set and 50 samples in PCDFs set. All these data are shown in Table 1 as observed values.

Table 1 Variables for TLSER equation, aqueous solubility and *n*-octanol/water partition coefficient

IUPAC	Substituted pattern	$V_{mc}/100$	π_1	ϵ_B	q_-	ϵ_A	q_+	$-\log S_w$			$\log K_{ow}$		
								Obs.	Pre.1	Pre.2	Obs.	Pre.1	Pre.2
Polychlorinated biphenyls (PCBs): 63 samples with $-\log S_w$, 154 samples with $\log K_{ow}$													
1	2-	1.72228	0.65234	0.13640	0.13329	0.12432	0.14701	4.54	4.54	4.79	4.38	4.72	4.57
2	3-	1.74772	0.66669	0.13414	0.13159	0.12095	0.14816	4.88	5.29	5.29	4.66	4.77	4.74
3	4-	1.74424	0.67994	0.13302	0.13102	0.12049	0.14712	5.20	5.70	5.54	4.63	4.81	4.80
4	2,2′-	1.77256	0.66623	0.13971	0.13212	0.12600	0.14769	5.27	5.15	5.53	4.90	5.15	4.91
5	2,3-	1.85124	0.64843	0.13812	0.13285	0.12271	0.15143				4.99	5.16	5.01
6	2,3′-	1.85064	0.64877	0.13794	0.13023	0.12262	0.15541				4.84	5.16	5.05
7	2,4-	1.84888	0.65931	0.13700	0.13263	0.12160	0.16200	5.25	5.51	5.63	5.15	5.17	5.10
8	2,4′-	1.85224	0.65544	0.13674	0.12982	0.12215	0.14885	5.28	5.46	5.60	5.09	5.16	5.06
9	2,5-	1.85244	0.65208	0.13743	0.13263	0.12195	0.15693	5.36	5.32	5.53	5.00	5.15	5.05
10	2,6-	1.82628	0.64842	0.14068	0.13281	0.12427	0.15105	5.21	4.97	5.52	4.93	5.21	4.97
11	3,3′-	1.87916	0.67159	0.13577	0.12269	0.11909	0.14942	5.80	6.25	6.26	5.30	5.32	5.32
12	3,4-	1.87608	0.67401	0.13434	0.13117	0.11873	0.15141	6.39	6.32	6.17	5.23	5.22	5.22
13	3,4′-	1.88204	0.67234	0.13463	0.12578	0.11867	0.14942				5.15	5.27	5.28
14	3,5-	1.88088	0.66235	0.13581	0.13117	0.11902	0.16310				5.40	5.21	5.22
15	4,4′-	1.87724	0.68670	0.13364	0.12514	0.11826	0.14842	6.56	6.75	6.49	5.23	5.31	5.36
16	2,3,2′-	1.89948	0.66258	0.14025	0.13124	0.12408	0.15161				5.60	5.53	5.32
17	2,4,2′-	1.90496	0.66740	0.14047	0.13114	0.12295	0.16266				5.60	5.58	5.42
18	2,5,2′-	1.90564	0.66414	0.13930	0.13124	0.12306	0.15821	6.02	5.91	6.12	5.55	5.51	5.36
19	2,6,2′-	1.87688	0.66612	0.13981	0.13211	0.12394	0.15169				5.60	5.46	5.26
20	2,3,3′-	1.9834	0.64420	0.13957	0.12420	0.12121	0.15596				5.60	5.64	5.53
21	2,3,4-	1.9822	0.65146	0.13829	0.13231	0.12018	0.15542				5.68	5.57	5.47
22	2,3,4′-	1.9852	0.65090	0.13822	0.12749	0.12077	0.15259	6.26	6.10	6.26	5.60	5.61	5.51
24	2,3,6-	1.95596	0.64847	0.14001	0.13233	0.12122	0.15650	6.29	5.82	6.18	5.50	5.55	5.40
25	2,4,3′-	1.98388	0.65301	0.13846	0.12545	0.12015	0.16332				5.54	5.63	5.59
26	2,5,3′-	1.98228	0.64883	0.13875	0.12392	0.12050	0.15717	6.01	6.01	6.23	5.65	5.62	5.56
28	2,4,4′-	1.98428	0.66058	0.13739	0.12728	0.11980	0.16316	6.21	6.38	6.39	5.71	5.63	5.61
29	2,4,5-	1.98264	0.65548	0.13780	0.13207	0.11948	0.16587	6.27	6.22	6.30	5.74	5.57	5.53
30	2,4,6-	1.95584	0.65307	0.14187	0.13224	0.12111	0.16547	6.14	5.91	6.41	5.50	5.67	5.50
31	2,5,4′-	1.98532	0.65476	0.13781	0.12720	0.12015	0.15719	6.25	6.23	6.32	5.70	5.61	5.55
32	2,6,4′-	1.9558	0.65014	0.13980	0.12828	0.12310	0.15228				5.53	5.60	5.43
33	3,4,2′-	1.98396	0.64997	0.13789	0.12929	0.12052	0.15969	6.29	6.04	6.15	5.71	5.57	5.50
34	3,5,2′-	1.982	0.64649	0.13952	0.12918	0.12081	0.16253				5.71	5.61	5.51
35	3,4,3′-	2.01336	0.66581	0.13586	0.12223	0.11705	0.15266				5.80	5.69	5.72
36	3,5,3′-	2.01268	0.65767	0.13739	0.12225	0.11729	0.16423				5.70	5.70	5.74
37	3,4,4′-	2.01048	0.67975	0.13487	0.12558	0.11668	0.15263	7.06	7.33	7.07	5.22	5.71	5.76
40	2,3,2′,3′-	2.03248	0.65779	0.14094	0.12387	0.12286	0.15259	7.28	6.46	6.72	5.67	5.97	5.80
41	2,3,4,2′-	2.03096	0.66323	0.14138	0.13079	0.12114	0.15612				6.00	5.97	5.78
42	2,3,2′,4′-	2.03916	0.66230	0.14118	0.12672	0.12195	0.16362				5.72	6.01	5.87
44	2,3,2′,5′-	2.03784	0.65950	0.14010	0.12386	0.12214	0.15880	6.47	6.54	6.70	6.00	5.96	5.84
45	2,3,6,2′-	2.00576	0.66513	0.14002	0.13169	0.12113	0.15681				4.84	5.84	5.68
46	2,3,2′,6′-	2.00604	0.66176	0.14018	0.12877	0.12293	0.15276				4.84	5.86	5.68
47	2,4,2′,4′-	2.03644	0.67012	0.14136	0.12704	0.12166	0.16370	6.51	6.86	7.08	5.94	6.06	5.92
48	2,4,5,2′-	2.0348	0.66550	0.14008	0.13080	0.12041	0.16652				5.69	5.93	5.83

Table 1(Cont'd)

IUPAC	Substituted pattern	$V_{mc}/100$	π_1	ε_B	q_-	ε_A	q_+	$-\log S_w$			$\log K_{ow}$		
								Obs.	Pre.1	Pre.2	Obs.	Pre.1	Pre.2
49	2,4,2′,5′-	2.09008	0.64951	0.14026	0.12717	0.12188	0.16381	6.57	6.52	6.66	6.10	6.03	5.92
50	2,4,6,2′-	2.00692	0.66921	0.14081	0.13159	0.12078	0.16606				5.75	5.90	5.77
51	2,4,2′,6′-	2.00616	0.66813	0.14071	0.12860	0.12275	0.16315				5.51	5.93	5.78
52	2,5,2′,5′-	2.09144	0.64599	0.14012	0.12065	0.12203	0.15877	7.00	6.44	6.60	6.17	6.04	5.95
53	2,5,2′,6′-	2.00712	0.66499	0.13931	0.12870	0.12281	0.15835	6.80	6.51	6.58	5.55	5.85	5.71
54	2,6,2′,6′-	1.92868	0.68411	0.14048	0.12949	0.12368	0.15219	7.21	6.65	6.84	5.90	5.80	5.60
55	2,3,4,3′-	2.11376	0.64805	0.13972	0.12364	0.11889	0.15655				6.10	6.06	5.98
60	2,3,4,4′-	2.11368	0.65570	0.13857	0.12716	0.11858	0.15658				6.24	6.04	5.98
61	2,3,4,5-	2.11156	0.65166	0.13913	0.13177	0.11813	0.16402	7.16	6.90	6.99	6.05	6.00	5.93
63	2,3,5,4′-	2.11472	0.65274	0.13925	0.12708	0.11876	0.16699				6.10	6.05	6.01
64	2,3,6,4′-	2.08712	0.65032	0.14065	0.12716	0.12019	0.15759				5.76	6.03	5.90
65	2,3,5,6-	2.08316	0.65025	0.14029	0.13188	0.11880	0.16953				5.96	5.95	5.87
66	2,4,3′,4′-	2.11268	0.65673	0.13848	0.12512	0.11845	0.16422	6.68	7.04	7.03	5.98	6.05	6.03
67	2,4,5,3′-	2.11324	0.65254	0.13910	0.12338	0.11819	0.16714				6.32	6.05	6.05
69	2,4,6,3′-	2.088	0.64973	0.14103	0.12611	0.11998	0.16665				6.03	6.05	5.95
70	2,5,3′,4′-	2.113	0.65138	0.13883	0.12003	0.11876	0.15985	7.25	6.89	6.95	6.22	6.06	6.04
71	2,6,3′,4′-	2.08736	0.64566	0.14017	0.12788	0.12215	0.15789				5.76	5.99	5.85
74	2,4,5,4′-	2.11416	0.65963	0.13821	0.12690	0.11793	0.16696				6.13	6.05	6.05
75	2,4,6,4′-	2.08944	0.65449	0.14078	0.12709	0.11996	0.16657	6.94	6.75	6.96	6.03	6.07	5.97
76	3,4,5,2′-	2.11412	0.64730	0.13908	0.12915	0.11880	0.16308				5.98	5.99	5.93
77	3,4,3′,4′-	2.13956	0.66154	0.13743	0.11718	0.11686	0.15634	8.53	7.46	7.36	6.21	6.15	6.19
80	3,5,3′,5′-	2.14376	0.64751	0.14053	0.11581	0.11736	0.16445	8.54	6.99	7.22	6.10	6.21	6.21
82	2,3,4,2′,3′-	2.1658	0.65882	0.14177	0.12294	0.12019	0.15715	7.05	7.34	7.55			
83	2,3,5,2′,3′-	2.16636	0.65763	0.14099	0.12296	0.12012	0.16695	6.96	7.26	7.36			
84	2,3,6,2′,3′-	2.13896	0.66019	0.14070	0.12386	0.12029	0.15772				6.10	6.28	6.17
85	2,3,4,2′,4′-	2.16292	0.66653	0.14214	0.12646	0.12006	0.16446				6.18	6.45	6.32
86	2,3,4,5,2′-	2.16496	0.66139	0.14119	0.13048	0.11883	0.16521	7.21	7.44	7.59	6.38	6.34	6.23
87	2,3,4,2′,5′-	2.16292	0.66289	0.14086	0.12059	0.12011	0.15912	7.91	7.44	7.54	6.54	6.40	6.31
88	2,3,4,6,2′-	2.13828	0.66626	0.14123	0.13128	0.11873	0.16944	7.43	7.42	7.57	6.50	6.29	6.19
89	2,3,4,2′,6′-	2.13844	0.66205	0.14178	0.12828	0.12094	0.15660				5.60	6.33	6.15
90	2,3,5,2′,4′-	2.1694	0.66332	0.14114	0.12653	0.11999	0.16704				6.32	6.40	6.31
91	2,3,6,2′,4′-	2.14108	0.66593	0.14095	0.12727	0.12015	0.16407				6.30	6.32	6.21
92	2,3,5,2′,5′-	2.16564	0.66113	0.14089	0.12016	0.12003	0.16707				6.50	6.40	6.34
93	2,3,5,6,2′-	2.1368	0.66456	0.14029	0.13128	0.11878	0.16980				6.06	6.23	6.15
95	2,3,6,2′,5′-	2.14088	0.66297	0.14005	0.12127	0.12020	0.15894				6.40	6.30	6.22
97	2,4,5,2′,3′-	2.16696	0.66171	0.14087	0.12296	0.11951	0.16746				6.60	6.39	6.32
98	2,4,6,2′,3′-	2.13944	0.66564	0.14107	0.12659	0.11982	0.16701				6.04	6.32	6.23
99	2,4,5,2′,4′-	2.17056	0.66745	0.14101	0.12652	0.11939	0.16751				6.54	6.42	6.34
100	2,4,6,2′,4′-	2.141	0.67075	0.14169	0.12719	0.11970	0.16707				6.23	6.39	6.27
101	2,4,5,2′,5′-	2.16408	0.66586	0.14088	0.12012	0.11941	0.16761	7.33	7.53	7.59	6.65	6.42	6.38
103	2,4,6,2′,5′-	2.1394	0.66796	0.14021	0.12654	0.11971	0.16716				6.11	6.30	6.23
104	2,4,6,2′,6′-	2.05828	0.68602	0.14158	0.12894	0.12055	0.16651	7.32	7.50	7.65			
105	2,3,4,3′,4′-	2.24508	0.65143	0.13963	0.11990	0.11741	0.16034				6.79	6.50	6.48
106	2,3,4,5,3′-	2.2438	0.64864	0.14042	0.12308	0.11696	0.16425				6.92	6.48	6.45
110	2,3,6,3′,4′-	2.2172	0.64685	0.14095	0.12044	0.11943	0.15847				6.20	6.45	6.37
112	2,3,5,6,3′-	2.21936	0.64607	0.14111	0.12344	0.11790	0.17052				6.41	6.43	6.39
113	2,3,6,3′,5′-	2.21868	0.64375	0.14164	0.12045	0.11936	0.16344				6.45	6.47	6.38
114	2,3,4,5,4′-	2.2438	0.65593	0.13941	0.12680	0.11673	0.16423				6.71	6.47	6.44
115	2,3,4,6,4′-	2.2178	0.65416	0.14143	0.12689	0.11788	0.16993				6.44	6.47	6.41
116	2,3,4,5,6-	2.21156	0.65235	0.14155	0.13154	0.11728	0.13953	7.92	7.65	7.97	6.55	6.41	6.20
117	2,3,5,6,4′-	2.219	0.65119	0.14122	0.12690	0.11782	0.17050				6.39	6.45	6.39
118	2,4,5,3′,4′-	2.24452	0.65587	0.13920	0.11966	0.11676	0.16799	7.39	7.81	7.72	6.61	6.50	6.53
119	2,4,6,3′,4′-	2.21768	0.65158	0.14107	0.12580	0.11907	0.16752				6.40	6.46	6.38
120	2,4,5,3′,5′-	2.24484	0.65077	0.14036	0.11696	0.11687	0.16813				6.30	6.53	6.55
121	2,4,6,3′,5′-	2.21884	0.64840	0.14205	0.12586	0.11894	0.16773				6.42	6.48	6.38
123	3,4,5,2′,4′-	2.24544	0.65352	0.13961	0.12447	0.11699	0.16507				6.64	6.48	6.46
128	2,3,4,2′,3′,4′-	2.2934	0.66247	0.14299	0.12027	0.11911	0.15799	9.01	8.21	8.37	6.96	6.89	6.76
129	2,3,4,5,2′,3′-	2.2968	0.65843	0.14193	0.12262	0.11795	0.16574	8.07	8.12	8.20	6.76	6.80	6.73

Table 1(Cont'd)

IUPAC	Substituted pattern	$V_{mc}/100$	π_1	ε_B	q_-	ε_A	q_+	$-\log S_w$			$\log K_{ow}$		
								Obs.	Pre.1	Pre.2	Obs.	Pre.1	Pre.2
130	2,3,4,2',3',5'-	2.29632	0.66016	0.14171	0.12032	0.11922	0.16769				7.30	6.83	6.77
131	2,3,4,6,2',3'-	2.26628	0.66371	0.14160	0.12343	0.11785	0.17036				6.78	6.72	6.67
132	2,3,4,2',3',6'-	2.27004	0.66160	0.14151	0.12108	0.11949	0.15853				6.20	6.75	6.65
134	2,3,5,6,2',3'-	2.26696	0.66116	0.14098	0.12344	0.11801	0.17061	8.60	8.01	8.00	7.30	6.68	6.64
135	2,3,5,2',3',6'-	2.27268	0.65993	0.14088	0.12075	0.11941	0.16709				7.30	6.72	6.67
136	2,3,6,2',3',6'-	2.18812	0.67976	0.14053	0.12165	0.12025	0.15799	8.65	8.09	8.02	6.96	6.59	6.49
137	2,3,4,5,2',4'-	2.29852	0.66468	0.14208	0.12634	0.11783	0.16549				7.03	6.84	6.74
138	2,3,4,2',4',5'-	2.29696	0.66449	0.14159	0.12030	0.11867	0.16828	8.32	8.25	8.23	7.00	6.85	6.80
139	2,3,4,6,2',4'-	2.26748	0.66993	0.14215	0.12704	0.11773	0.17041				6.58	6.78	6.69
141	2,3,4,5,2',5'-	2.29376	0.66236	0.14149	0.11976	0.11785	0.16568	7.68	8.23	8.25	6.75	6.82	6.77
143	2,3,4,5,2',6'-	2.26656	0.66201	0.14119	0.12794	0.11867	0.16532				6.56	6.68	6.58
144	2,3,4,6,2',5'-	2.26632	0.66684	0.14077	0.12046	0.11775	0.17050				6.45	6.73	6.71
146	2,3,5,2',4',5'-	2.29764	0.66399	0.14163	0.11782	0.11857	0.16849				6.85	6.87	6.83
149	2,3,6,2',4',5'-	2.27232	0.66414	0.14084	0.12079	0.11935	0.16790				6.80	6.75	6.70
151	2,3,5,6,2',5'-	2.26984	0.66348	0.14071	0.12049	0.11793	0.17073	7.42	8.10	8.04	6.42	6.71	6.70
153	2,4,5,2',4',5'-	2.29788	0.66732	0.14162	0.11842	0.11854	0.16842	8.56	8.35	8.32	7.09	6.89	6.85
154	2,4,5,2',4',6'-	2.2718	0.66909	0.14101	0.12627	0.11885	0.16796				6.65	6.75	6.68
155	2,4,6,2',4',6'-	2.19192	0.68773	0.14361	0.12686	0.11953	0.16745	8.71	8.34	8.52	7.07	6.75	6.60
156	2,3,4,5,3',4'-	2.37412	0.65255	0.14037	0.11953	0.11569	0.16444	7.82	8.50	8.41	7.44	6.92	6.93
158	2,3,4,6,3',4'-	2.35004	0.65058	0.14166	0.11986	0.11708	0.17084	7.66	8.19	8.22	6.78	6.91	6.88
163	2,3,5,6,3',4'-	2.35076	0.64761	0.14162	0.11890	0.11713	0.17128				6.78	6.89	6.87
165	2,3,5,6,3',5'-	2.34772	0.64567	0.14186	0.11041	0.11708	0.17142				7.00	6.93	6.95
167	2,4,5,3',4',5'-	2.37572	0.65320	0.14024	0.11666	0.11551	0.16880				7.29	6.94	6.98
168	2,4,6,3',4',5'-	2.3478	0.64959	0.14264	0.12564	0.11824	0.16839				6.63	6.92	6.81
169	3,4,5,3',4',5'-	2.40160	0.66750	0.13831	0.10895	0.11241	0.15571				7.55	7.05	7.17
170	2,3,4,5,2',3',4'-	2.42736	0.66158	0.14265	0.12011	0.11714	0.16600				7.08	7.27	7.20
171	2,3,4,6,2',3',4'-	2.40008	0.66514	0.14273	0.12081	0.11704	0.17113	8.30	8.91	8.89	7.11	7.21	7.16
172	2,3,4,5,2',3',5'-	2.42872	0.65960	0.14228	0.10591	0.11705	0.16841				7.21	7.33	7.35
174	2,3,4,5,2',3',6'-	2.39748	0.66203	0.14183	0.12039	0.11782	0.16583				6.85	7.15	7.09
175	2,3,4,6,2',3',5'-	2.40052	0.66365	0.14157	0.12011	0.11693	0.17134				6.92	7.15	7.13
176	2,3,4,6,2',3',6'-	2.32340	0.67969	0.14136	0.12111	0.11773	0.17074				6.55	7.02	6.99
177	2,3,5,6,2',3',4'-	2.40172	0.66201	0.14170	0.12083	0.11727	0.17133				6.73	7.15	7.12
178	2,3,5,6,2',3',5'-	2.40264	0.66057	0.14146	0.10421	0.11723	0.17147				6.85	7.23	7.29
179	2,3,5,6,2',3',6'-	2.32176	0.67786	0.14084	0.12115	0.11812	0.17072				7.00	6.98	6.96
180	2,3,4,5,2',4',5'-	2.42796	0.66433	0.14221	0.11756	0.11706	0.16910				7.21	7.29	7.26
181	2,3,4,5,6,2',4'-	2.40156	0.66775	0.14238	0.12711	0.11623	0.16553				7.13	7.17	7.08
182	2,3,4,5,2',4',6'-	2.39948	0.66630	0.14203	0.12615	0.11773	0.16865				6.92	7.16	7.08
183	2,3,4,6,2',4',5'-	2.40168	0.66769	0.14154	0.12005	0.11694	0.17126	7.92	9.00	8.85	7.04	7.19	7.17
185	2,3,4,5,6,2',5'-	2.39916	0.66488	0.14129	0.12013	0.11625	0.15978	8.46	9.01	8.91	7.00	7.14	7.09
186	2,3,4,5,6,2',6'-	2.31444	0.68111	0.14145	0.12837	0.11704	0.15477				6.78	6.96	6.83
187	2,3,5,6,2',4',5'-	2.40032	0.66541	0.14149	0.11819	0.11720	0.17144	8.94	8.91	8.76			
189	2,3,4,5,3',4',5'-	2.50924	0.64957	0.14140	0.10094	0.11455	0.16538				7.72	7.47	7.56
190	2,3,4,5,6,3',4'-	2.47716	0.65165	0.14219	0.11953	0.11563	0.15931				7.08	7.32	7.26
191	2,3,4,6,3',4',5'-	2.48128	0.64863	0.14320	0.11938	0.11633	0.17168				7.21	7.37	7.32
192	2,3,4,5,6,3',5'-	2.47552	0.64896	0.14297	0.10957	0.11552	0.16474				7.21	7.39	7.38
193	2,3,5,6,3',4',5'-	2.47688	0.64696	0.14243	0.10253	0.11649	0.17201				7.21	7.40	7.47
194	2,3,4,5,2',3',4',5'-	2.55468	0.66268	0.14317	0.09041	0.11643	0.16641	9.16	9.75	9.60	8.15	7.87	7.95
195	2,3,4,5,6,2',3',4'-	2.52864	0.66498	0.14289	0.12062	0.11555	0.15904				7.35	7.61	7.52
196	2,3,4,5,2',3',4',6'-	2.52976	0.66518	0.14254	0.11853	0.11629	0.17200				7.43	7.62	7.59
202	2,3,5,6,2',3',5',6'-	2.44740	0.67867	0.14132	0.09818	0.11739	0.17143	9.15	9.56	9.42	8.28	7.53	7.63
203	2,3,4,5,6,2',4',5'-	2.53028	0.66738	0.14207	0.12082	0.11545	0.15948				7.49	7.59	7.53
204	2,3,4,5,6,2',4',6'-	2.44880	0.68379	0.14227	0.12660	0.11610	0.16882				7.48	7.44	7.37
205	2,3,4,5,6,3',4',5'-	2.60716	0.65024	0.14362	0.08975	0.11487	0.16263				7.62	7.94	8.03
206	2,3,4,5,6,2',3',4',5'-	2.66100	0.66455	0.14302	0.08357	0.11482	0.16653	10.26	10.48	10.20	8.20	8.23	8.38
207	2,3,4,5,6,2',3',4',6'-	2.58336	0.68063	0.14270	0.12033	0.11544	0.17209				8.20	7.88	7.86
208	2,3,4,5,6,2',3',5',6'-	2.57528	0.68004	0.14198	0.09739	0.11542	0.17194	10.41	10.41	10.02	8.20	7.95	8.07

Table 1(Cont'd)

IUPAC	Substituted pattern	$V_{mc}/100$	π_1	ε_B	q_-	ε_A	q_+	$-\log S_w$			$\log K_{ow}$		
								Obs.	Pre.1	Pre.2	Obs.	Pre.1	Pre.2
209	2,3,4,5,6,2',3',4',5',6'-	2.70848	0.68071	0.14301	0.06297	0.11484	0.05247	11.62	11.88	11.87	8.58	8.61	8.43
Polychlorinated dibenzo-*p*-dioxins (PCDDs): 14 samples with $-\log S_w$, 42 samples with $\log K_{ow}$													
1	1-	1.81908	0.68675	0.13166	0.13262	0.12238	0.15830	5.72	6.16	5.32	5.05	5.05	5.15
2	2-	1.81928	0.69577	0.13154	0.13167	0.12165	0.16928	5.87	6.41	5.82	5.00	5.11	5.19
10	2,3-	1.96936	0.68857	0.13251	0.12879	0.11952	0.17335	7.23	7.10	6.80	5.60	5.56	5.65
11	2,7-	1.94820	0.69827	0.13275	0.12856	0.11971	0.17026	7.83	7.29	7.53	5.75	5.58	5.54
12	2,8-	1.97620	0.68808	0.13271	0.12899	0.11968	0.17062	7.18	7.13	6.95	5.60	5.84	5.72
14	1,2,4-	2.08284	0.68324	0.13373	0.12418	0.11839	0.17095	7.58	7.64	7.53	6.35	6.06	5.97
27	1,2,3,4-	2.23192	0.67759	0.13435	0.12325	0.11682	0.15969	8.77	8.44	8.46	6.48	6.25	6.12
28	1,2,3,6-	2.23624	0.67465	0.13472	0.12291	0.11707	0.17892				6.86	6.46	6.25
29	1,2,3,7-	2.23096	0.68533	0.13453	0.12428	0.11661	0.17742	8.89	8.58	9.14	6.48	6.66	6.33
30	1,2,3,8-	2.23728	0.68272	0.13455	0.12468	0.11662	0.17706				6.10	6.44	6.38
31	1,2,3,9-	2.24008	0.67218	0.13465	0.12579	0.11712	0.17668				6.39	6.37	6.34
32	1,2,4,6-	2.21420	0.67334	0.13576	0.12318	0.11748	0.17179				6.10	6.37	6.12
33	1,2,4,7-	2.23740	0.67870	0.13482	0.11942	0.11703	0.17331				6.25	6.46	6.47
34	1,2,4,8-	2.23808	0.67827	0.13476	0.11930	0.11695	0.17349				6.25	6.45	6.49
35	1,2,4,9-	2.21096	0.67673	0.13505	0.12345	0.11749	0.17167				6.10	6.35	6.23
37	1,2,6,8-	2.24128	0.67491	0.13488	0.12119	0.11716	0.17623				6.43	6.44	6.43
40	1,2,7,9-	2.23960	0.67433	0.13487	0.12561	0.11727	0.17387				6.86	6.40	6.30
42	1,3,6,8-	2.24020	0.67673	0.13515	0.12127	0.11701	0.17613	9.00	8.36	8.65	6.29	6.45	6.37
43	1,3,6,9-	2.21188	0.67827	0.13522	0.12171	0.11760	0.17581				6.25	6.38	6.25
44	1,3,7,8-	2.24192	0.68294	0.13475	0.12419	0.11661	0.17687				6.30	6.47	6.37
45	1,3,7,9-	2.23128	0.67923	0.13498	0.12495	0.11696	0.17432				6.39	6.42	6.27
46	1,4,6,9-	2.22244	0.66357	0.13666	0.11609	0.11850	0.15955				6.38	6.42	6.18
47	1,4,7,8-	2.23028	0.67955	0.13484	0.11588	0.11719	0.17659				6.39	6.46	6.54
48	2,3,7,8-	2.22848	0.69391	0.13450	0.12351	0.11627	0.17535	10.22	8.85	9.69	6.42	6.49	6.39
49	1,2,3,4,6-	2.36576	0.67074	0.13556	0.12259	0.11586	0.16333				6.30	6.80	6.72
50	1,2,3,4,7-	2.36748	0.67975	0.13537	0.11865	0.11540	0.17403	9.48	9.25	9.40	6.60	6.88	6.87
51	1,2,3,6,7-	2.36816	0.67717	0.13548	0.11372	0.11569	0.17975				6.74	6.90	6.99
52	1,2,3,6,8-	2.35432	0.68241	0.13569	0.12052	0.11550	0.17965				6.53	6.87	6.72
53	1,2,3,6,9-	2.34232	0.67836	0.13578	0.11551	0.11605	0.17933				6.24	6.84	6.79
54	1,2,3,7,8-	2.37064	0.68443	0.13535	0.12207	0.11519	0.17810				6.64	6.90	6.79
55	1,2,3,7,9-	2.37196	0.67502	0.13597	0.12287	0.11558	0.17778				6.40	6.87	6.67
59	1,2,4,6,9-	2.35620	0.66481	0.13731	0.11527	0.11658	0.17238				6.60	6.86	6.58
60	1,2,4,7,8-	2.36148	0.68230	0.13558	0.11460	0.11556	0.17741				6.20	6.92	6.92
64	1,2,3,4,6,8-	2.50348	0.67338	0.13646	0.12031	0.11441	0.17713				6.85	7.29	7.13
66	1,2,3,4,7,8-	2.50192	0.68104	0.13613	0.11126	0.11411	0.17790	9.95	10.07	9.71	7.80	7.38	7.43
68	1,2,3,6,7,9-	2.50248	0.67354	0.13646	0.11435	0.11458	0.18012				7.59	7.33	7.29
69	1,2,3,6,8,9-	2.50284	0.67288	0.13649	0.11387	0.11460	0.17994				7.59	7.33	7.30
71	1,2,4,6,7,9-	2.50800	0.66189	0.13801	0.11417	0.11520	0.17322				6.85	7.33	7.04
72	1,2,4,6,8,9-	2.50212	0.66323	0.13789	0.11375	0.11508	0.17317				6.85	7.32	7.05
73	1,2,3,4,6,7,8-	2.63336	0.67506	0.13698	0.11265	0.11325	0.16813	11.25	10.72	10.83	8.00	7.77	7.72
75	1,2,3,4,6,7,8,9-	2.71108	0.67931	0.13893	0.09986	0.11257	0.06445	12.79	11.90	12.60	8.20	8.19	8.00
Polychlorinated dibenzofurans (PCDFs): 7 samples with $-\log S_w$, 50 samples with $\log K_{ow}$													
16	2,8-	1.85808	0.70513	0.13683	0.12046	0.11753	0.15970	7.21	7.18	7.21	5.44	5.54	5.56
49	1,2,3,4-	2.11808	0.69454	0.13720	0.14977	0.11406	0.15806				6.17	6.09	6.07
50	1,2,3,6-	2.11612	0.69419	0.13762	0.14160	0.11420	0.17682				6.15	6.15	6.09
52	1,2,3,8-	2.12000	0.69605	0.13766	0.11965	0.11404	0.17577				6.15	6.30	6.21
53	1,2,3,9-	2.11052	0.69085	0.13703	0.15129	0.11423	0.17569				6.06	6.03	6.01
56	1,2,4,8-	2.25624	0.64835	0.13721	0.11842	0.11428	0.17015				6.31	6.37	6.14
58	1,2,6,7-	2.12172	0.69289	0.13755	0.13924	0.11428	0.16525				6.25	6.17	6.11
61	1,2,7,8-	2.12216	0.69501	0.13755	0.11950	0.11416	0.17325				6.23	6.30	6.20
62	1,2,7,9-	2.13820	0.68979	0.13755	0.11950	0.11416	0.17325				6.25	6.31	6.20
64	1,3,4,6-	2.12156	0.69087	0.13745	0.14157	0.11402	0.16862				6.31	6.13	6.08
65	1,3,4,7-	2.11996	0.70549	0.13660	0.14587	0.11353	0.17153				6.23	6.17	6.19
66	1,3,4,8-	2.12160	0.69592	0.13693	0.13689	0.11382	0.16901				6.13	6.17	6.14
67	1,3,4,9-	2.11452	0.68966	0.13692	[illegible]3089	0.11406	0.16851				5.89	6.03	6.01
69	1,3,6,8-	2.12308	0.69475	0.13804	0.14367	0.11389	0.17450				6.37	6.18	6.11

Table 1(Cont'd)

	IUPAC Substituted pattern	$V_{mc}/100$	π_1	ε_B	q_-	ε_A	q_+	$-\log S_w$			$\log K_{ow}$		
								Obs.	Pre.1	Pre.2	Obs.	Pre.1	Pre.2
71	1,3,7,8-	2.12208	0.70435	0.13686	0.14319	0.11358	0.17352				6.34	6.19	6.20
72	1,3,7,9-	2.14100	0.69812	0.13686	0.14319	0.11358	0.17352				6.34	6.21	6.19
73	1,4,6,7-	2.11820	0.69170	0.13731	0.13946	0.11417	0.15783				6.15	6.14*	6.08
75	1,4,6,9-	2.11300	0.68226	0.13723	0.14400	0.11446	0.15806				5.60	6.03	5.96
77	2,3,4,6-	2.12104	0.69727	0.13772	0.14045	0.11426	0.16253				6.11	6.20	6.14
78	2,3,4,7-	2.11880	0.71067	0.13727	0.14499	0.11376	0.17193				6.06	6.24	6.24
80	2,3,4,9-	2.12132	0.69094	0.13743	0.14502	0.11421	0.17046				6.17	6.12	6.06
81	2,3,6,7-	2.11908	0.71030	0.13701	0.13783	0.11381	0.17479				6.31	6.27	6.27
82	2,3,6,8-	2.11868	0.70500	0.13817	0.11643	0.11401	0.17489				6.73	6.40	6.30
83	2,3,7,8-	2.11552	0.71529	0.13738	0.11758	0.11370	0.17393	8.86	9.02	8.77	6.53	6.42	6.38
84	2,4,6,7-	2.11908	0.70035	0.13793	0.13821	0.11415	0.16737				6.25	6.24	6.17
85	2,4,6,8-	2.12072	0.69545	0.13781	0.09643	0.11437	0.16757				6.17	6.44	6.29
88	1,2,3,4,7-	2.24712	0.70266	0.13767	0.14554	0.11229	0.17235				6.53	6.58	6.52
89	1,2,3,4,8-	2.24800	0.69443	0.13790	0.11732	0.11255	0.16529				6.79	6.70	6.55
91	1,2,3,6,7-	2.24932	0.69956	0.13779	0.13836	0.11249	0.17780				6.26	6.61	6.52
92	1,2,3,6,8-	2.24488	0.69477	0.13897	0.11743	0.11265	0.17788				6.33	6.74	6.55
93	1,2,3,6,9-	2.24440	0.68850	0.13756	0.14318	0.11276	0.17794				6.65	6.48	6.39
94	1,2,3,7,8-	2.24832	0.70277	0.13788	0.11855	0.11236	0.17689				6.79	6.75	6.63
97	1,2,4,6,7-	2.24632	0.69276	0.13863	0.13879	0.11276	0.17106				6.27	6.59	6.45
98	1,2,4,6,8-	2.25240	0.68576	0.13801	0.09447	0.11293	0.17123				6.34	6.79	6.58
99	1,2,4,6,9-	2.24504	0.68135	0.13760	0.14365	0.11306	0.17139				6.59	6.44	6.32
100	1,2,4,7,8-	2.25260	0.69442	0.13803	0.11712	0.11260	0.17531				6.26	6.72	6.57
101	1,2,4,7,9-	2.24388	0.69109	0.13844	0.14987	0.11260	0.17541				6.19	6.50	6.39
103	1,3,4,6,7-	2.24868	0.69762	0.13763	0.13834	0.11237	0.16957				6.19	6.59	6.50
104	1,3,4,6,8-	2.24996	0.69213	0.13813	0.13639	0.11250	0.16994				6.24	6.59	6.46
105	1,3,4,6,9-	2.24524	0.68690	0.13799	0.14304	0.11267	0.16943				6.34	6.49	6.38
107	1,3,4,7,9-	2.24768	0.69655	0.13741	0.14890	0.11223	0.17538				6.33	6.51	6.45
108	1,3,4,8,9-	2.24400	0.68938	0.13768	0.14319	0.11271	0.16962				6.51	6.49	6.40
109	2,3,4,6,7-	2.24680	0.70460	0.13807	0.13737	0.11254	0.16319				6.47	6.66	6.57
110	2,3,4,6,8-	2.25000	0.69888	0.13882	0.09553	0.11269	0.16836				6.59	6.90	6.69
111	2,3,4,6,9-	2.24992	0.68960	0.13784	0.13648	0.11275	0.17083				6.53	6.56	6.44
112	2,3,4,7,8-	2.24876	0.70896	0.13826	0.11545	0.11238	0.17572	9.16	9.63	9.57	6.92	6.83	6.70
114	2,3,4,8,9-	2.24944	0.69153	0.13873	0.11721	0.11284	0.17124				6.42	6.72	6.53
118	1,2,3,4,7,8-	2.37772	0.70069	0.13840	0.11617	0.11100	0.17610	10.66	10.16	10.35	7.00	7.16	6.98
121	1,2,3,6,7,8-	2.38016	0.69793	0.13896	0.11650	0.11115	0.17868	10.33	10.06	10.36			
131	1,2,3,4,6,7,8-	2.50756	0.69601	0.13941	0.07686	0.10992	0.17183	11.48	10.81	11.13	7.40	7.79	7.45
135	1,2,3,4,6,7,8,9-	2.63008	0.69242	0.14001	0.08270	0.10886	0.06183	11.58	12.06	11.87	7.97	8.13	7.73

For every compound in each sample set, five quantum chemical parameters were calculated individually using WinMOPAC(version 2.0) software(Stewart, 1990): the average molecular polarizability(α), the energy of the lowest unoccupied molecular orbital(E_{LUMO}), the energy of the highest occupied molecular orbital(E_{HOMO}), the most positive charge on a hydrogen atom(q_+), and the most negative atomic partial charge (q_-) in the solute molecule. AM1 algorithm was selected and MOPAC was run with the following keywords string: "AM1 EF PI SCFCRT = 1.D—12 PRECISE GEO-OK LET BONDS LARGE SUPER MULLIK GRAPH POLAR".

TLSER methodology, as a theoretical extension of the empirical Kamlet-Taft-Abraham linear solvation energy relationship (LSER) method, uses a single set of descriptors and each of them describes a single, independent molecular event or characteristics:

$$P = P_0 + a \cdot V_{mc}/100 + b \cdot \pi_1 + c \cdot \varepsilon_B + d \cdot q_- + e \cdot \varepsilon_A + f \cdot q_+ . \quad (1)$$

Where P is the interested property; P_0 is the constant term; V_{mc} is the molecular van der Waals volume, in cubic angstroms, and the polarizability term π_1 is derived as the ratio of the molecular polarizability (α) and V_{mc}. The covalent contribution to Lewis basicity, ε_B, is calculated as the difference in energy between the lowest unoccupied molecular orbital (E_{LUMO}) of standard solvent(water) and the highest occupied molecular orbital(E_{HOMO}) of solute, i.e. $|E_{HOMO}-E_{LUMO,Water}|$, with energy units of electron volt(eV). The electrostatic contribution to basicity, q_-, is simply the most negative atomic partial charge in the solute molecule. Analogously, the hydrogen-bonding donating ability is divided into two components: The covalent contribution to Lewis acidity, ε_A, is the energy difference between the E_{HOMO} of water and E_{LUMO} of solute, i.e. $|E_{LUMO}-E_{HOMO,Water}|$, with energy units of electron volt (eV). The electrostatic contribution to Lewis acidity of the solvent, q_+, is the most positive formal charge on a hydrogen atom in the solute molecule. Six symbols, a, b, c, d, e, f, are all constant terms calculated using multiple linear regression (MLR) method (Wilson, 1991).

For those independent variables in Eq. (1), V_{mc} was calculated using the software on the CD-ROM with Karelson's book, named

"Moldes"(Karelson, 2000), π_1 was then derived from α/V_{mc}. For the calculation of ε_B and ε_A, $E_{HOMO} = -12.536$ eV and $E_{LUMO} = 4.452$ eV were used for the water (Karelson, 2000). After all variables were ready, forward stepwise multiple linear regression (MLR) was made between aqueous solubility or *n*-octanol/water partition coefficient and the quantum chemical descriptors using MINITAB (Release 13.1) software(MINITAB Inc., 2002). To reflect the difference among three chemical groups, MLR equations were established both for their combination as well as for PCBs, PCDDs and PCDFs respectively. The default options were adopted for MLR process in MINITAB: Alpha-to-Enter = 0.05.

2 Results and discussion

The TLSER independent variables, calculated from the parameters given by Moldes and WinMOPAC, are shown in Table 1.

2.1 Predicting the aqueous solubility ($-\log S_w$)

The regression equation for the whole sample set containing 84 chemicals with aqueous solubility(63 PCBs, 14 PCDDs and 7 PCDFs) is:

$$-\log S_w = -27.260 + 6.03\ V_{mc}/100 + 34.2\ \pi_1 - 5.7\ q_+,\ n = 84,\ R^2 = 0.9230, \quad (2)$$
$$R^2_{adj} = 0.9201, S = 0.489, F = 319.69, p = 0.000.$$

Regression equations for PCBs, PCDDs and PCDFs subsets are shown below respectively:

(1) For 63 PCBs subset:

$$-\log S_w = -18.834 + 5.61\ V_{mc}/100 + 23.2\ \pi_1 - 7.7\ q_+,$$
$$n = 63,\ R^2 = 0.9046,\ R^2_{adj} = 0.8997, \quad (3)$$
$$S = 0.443,\ F = 186.45,\ p = 0.000.$$

(2) For 14 PCDDs subset:

$$-\log S_w = -283.3 + 1670\ \varepsilon_B + 74\ \pi_1 + 136\ q_-,$$
$$n = 14,\ R^2 = 0.9800,\ R^2_{adj} = 0.9739, \quad (4)$$
$$S = 0.320,\ F = 162.96,\ p = 0.000$$

(3) For 7 PCDFs subset:

$$-\log S_w = -4.012 + 6.04\ V_{mc}/100, \quad (5)$$
$$n = 7,\ R^2 = 0.9678,\ R^2_{adj} = 0.9614,$$
$$S = 0.310,\ F = 150.30,\ p = 0.000.$$

The predicted values calculated by Eq.(2) are given in Table 1 as Pre.1 $-\log S_w$, and those calculated by Eqs.(3)—(5) are listed in Table 1 as Pre.2 $-\log S_w$. The relationship between predicted values and observed values are shown in Fig. 1.

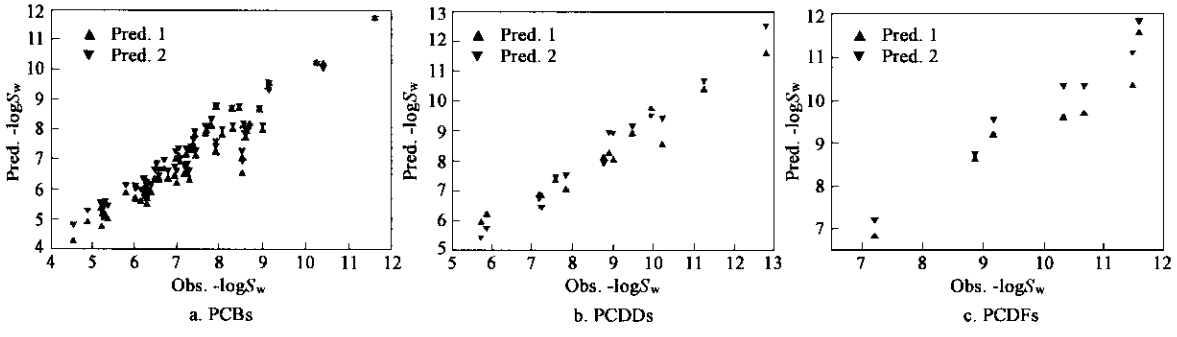

Fig.1 Predicted $-\log S_w$ value VS. observed $-\log S_w$ values for PCBs, PCDDs and PCDGs

The predicted values are quite close to the observed values, both from Eq. (2) and from Eqs. (3)—(5). The predicted values from Eq. (2) and values from Eqs. (3)—(5) are in fact quite similar, though the latter seems a bit better for some samples in PCDDs and PCDFs subset. In addition, judged from the performance parameters such like high R^2_{adj}, which seems quite low S, large F and small p, Eqs. (2)—(5) are all of significant related and quite satisfying accuracy.

In order to find the relative importance of variables for the QSPR express, it was necessary and convenient to analyze the level and PN sign of each independent variable in regression equations. The t parameter and p-level for each variable in Eqs.(2)—(5) are shown in Table 2.

Table 2 t parameter and p-level of each independent invariables in Eqs. (2)—(5)

Variable or constant	t parameter				p-level			
	Eq.(2)	Eq.(3)	Eq.(4)	Eq.(5)	Eq.(2)	Eq.(3)	Eq.(4)	Eq.(5)
$V_{mc}/100$	26.43	22.18		12.26	0.000	0.000		0.000
π_1	10.32	4.35	4.33		0.000	0.000	0.001	
ε_B			7.84				0.000	
q_-			2.95				0.015	
ε_A								
q_+	-2.21	-2.10			0.030	0.040		

All variables included in final equations are significant at the level of $p \leqslant 0.05$. $V_{mc}/100$ is the most important variable in all equations except Eq.(4), as it occupied the largest corresponding t-parameter, and the $-\log S_w$ increases with the $V_{mc}/100$'s increase, which means the aqueous solubility could decrease. The reason is that $V_{mc}/100$ can reflect the adsorbed energy during the formation of cavity in solvent. The H_2O molecular has large polarity and the combination energy is quite large, so the solute with larger volume (V_{mc}) does not tend to be dissolved by solvent(water).

According to the assumption of TLSER model, a specific property was the sum of three terms: steric term (represented by $V_{mc}/100$), polarizability term(represented by π_1) and hydrogen-bonding term which was further divided into donor and acceptor components. ε_B and q_- respectively reflected the covalent and electrostatic contribution to donor component, while ε_A and q_+ respectively reflected the covalent and electrostatic contribution to acceptor component. As shown in Table 2, the steric term and polarizability term were predominant for all equations except Eq. (4); hydrogen-bonding term had relatively lower effect on the aqueous solubility. The exception of Eq.(4) is somewhat caused by the small size and abnormal sample distribution of PCDDs subset, where most samples are have low chlorinated.

2.2 Predicting the *n*-octanol/water partition coefficient ($\log K_{ow}$)

The regression equation for the whole sample set containing 246 chemicals with *n*-octanol/water partition coefficients (154 PCBs, 42 PCDDs and 50 PCDFs) is:

$$\log K_{ow} = -10.215 + 3.012\ V_{mc}/100 + 46.4\ \varepsilon_B + 6.4\pi_1 - 6.1 q_-,$$
$$n = 246,\ R^2 = 0.8779,\ R^2_{adj} = 0.8759,\ S = 0.255,\ F = 383.79,\ p = 0.000. \quad (6)$$

Regression equations for PCBs, PCDDs and PCDFs subsets are shown below respectively:

(1) For 154 PCBs subset:

$$\log K_{ow} = -4.617 + 3.498\ V_{mc}/100 + 6.3\ \pi_1 - 6.6\ q_-,$$
$$n = 154,\ R^2 = 0.9093,\ R^2_{adj} = 0.9074,\ S = 0.249,\ F = 501.03,\ p = 0.000. \quad (7)$$

(2) For 42 PCDDs subset:

$$\log K_{ow} = -55.9463 + 6.37\ V_{mc}/100 + 28.3\ \pi_1 + 246\ \varepsilon_A,$$
$$n = 42,\ R^2 = 0.8569,\ R^2_{adj} = 0.8456,\ S = 0.255,\ F = 75.84,\ p = 0.000. \quad (8)$$

(3) For 50 PCDFs subset:

$$\log K_{ow} = -5.6730 + 2.77\ V_{mc}/100 + 9.3\ \pi_1 - 3.9\ q_-,$$
$$n = 50,\ R^2 = 0.7937,\ R^2_{adj} = 0.7802,\ S = 0.184,\ F = 58.97,\ p = 0.000. \quad (9)$$

The predicted values calculated by Eq.(6) are given in Table 1 as Pre.1 $\log K_{ow}$, and those calculated by Eqs.(7)—(9) are listed in Table 1 as Pre.2 $\log K_{ow}$. The relationship between predicted values and observed values are shown in Fig.2.

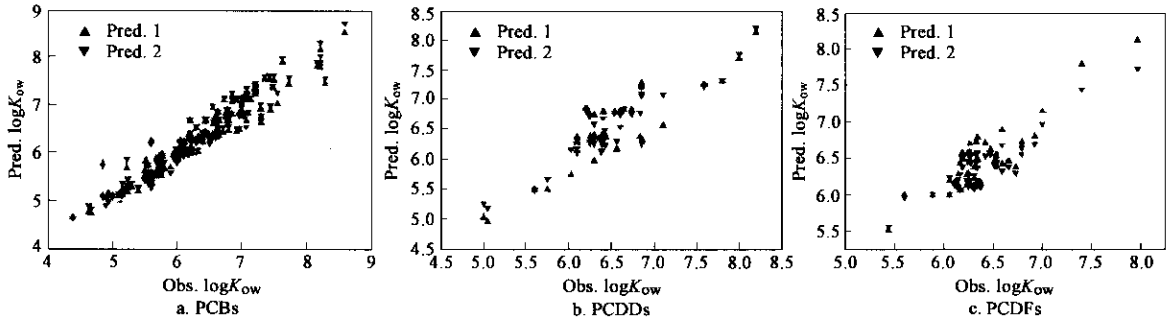

Fig.2 Predicted $\log K_{ow}$ value VS. observed $\log K_{ow}$ values for 254 PCBs, PCDDs and PCDFs

The predicted values is quite close to the observed values, both from Eq. (6) and from Eqs. (7)—(9) which are in fact quite similar, though the latter seems a bit better for some samples in PCDDs and PCDFs subset. In addition, judged from the performance parameters such like high R^2_{adj}, which seems quite low S, large F and small p, Eqs.(6)—(9) are all of significant related and quite satisfying accuracy.

Like the analysis of variables in $-\log S_w$ regression equations, the t parameter and p-level for each variable in Eqs.(6)—(9) are listed in Table 3.

Table 3 t parameter and p-level of each independent invariables in Eqs. (6)—(9)

Variable or constant	t parameter				p-level			
	Eq.(6)	Eq.(7)	Eq.(8)	Eq.(9)	Eq.(6)	Eq.(7)	Eq.(8)	Eq.(9)
$V_{mc}/100$	26.11	25.07	4.96	10.20	0.000	0.000	0.000	0.000
π_1	5.54	3.08	4.99	3.40	0.000	0.002	0.000	0.001
ε_B	5.77				0.000			
q_-	-3.59	-2.23			0.000	0.027	0.036	0.024
ε_A			2.18	-2.33				
q_+								

All variables included in final equations are significant at the level of $p \leqslant 0.05$. $V_{mc}/100$ is the most important variable in all equations except Eq.(8), as it occupied the largest corresponding t-parameter. In Eq.(8), the t-parameter of $V_{mc}/100$ is quite close to π_1. The $\log K_{ow}$ increases with the $V_{mc}/100$'s increase, which means the hydrophobicity would decrease. The reason is quite similar with the discussion about aqueous solubility above. As $V_{mc}/100$ can reflect the adsorbed energy during the formation of cavity in solvent, the H_2O molecular has large polarity and the combination energy is quite large, so the solute with larger volume(V_{mc}) would tend to stay in organic phase.

2.3 Robustness of the models

It is very important for the equations developed in QSAR/QSPR work to be used in the prediction of property values for the compounds not employed in the development of those equations. So it is necessary to obtain a measure of the model predictive capacity and stability, i.e. robustness. A systematic algorithm for the estimation of predictive power of the regression equation is given by the PRESS statistic (Karelson, 2000). And the most usual way to explore the stability of a predictive model is through the analysis of the influence of each one of the individual objects that configure the final model, which is known as cross-validation by leave-one-out(Carbo, 2000). PRESS(prediction sum of squares) is defined as the sum of residues from leave-one-out method, from which the cross-validated coefficient of the determination of the multiple regression, and RMSPE(root-mean-square prediction error) is the normalized form of PRESS. R^2_{pred} (i.e. R^2_{CV}) could then be calculated from PRESS value. R^2_{pred} is expected to be close to the coefficient of determination, R^2, for the multiple regressions involving the whole data set. For the choice of the best multiple regression model, the model with the smallest PRESS value may be preferred(Karelson, 2000). And a value $R^2_{pred} > 0.5$ is commonly accepted as satisfactory results(Carbo, 2000).

MINITAB can perform the cross-validation operation; the results were also given in Table 4. Obviously, all RMSPE values keep at a level

less than 0.6, which suggests good predictive capacity of the developed equations. And all R^2_{pred} values are much greater than 0.5 and quite close to the R^2 values, which suggests good stability of them.

Table 4 Robustness parameters of the equations

Variable or constant	Equations for $-\log S_w$				Equations for $\log K_{ow}$			
	Eq.(2)	Eq.(3)	Eq.(4)	Eq.(5)	Eq.(6)	Eq.(7)	Eq.(8)	Eq.(9)
PRESS	21.6274	14.5394	2.1573	0.9503	16.2819	9.8133	2.8933	1.9183
RMSPE	0.5074	0.4804	0.3925	0.3485	0.2573	0.2524	0.2625	0.1959
R^2_{pred}	0.9128	0.8803	0.9579	0.9431	0.8728	0.9040	0.8328	0.7467
R^2	0.9230	0.9046	0.9800	0.9678	0.8779	0.9093	0.8569	0.7937

3 Conclusions

TLSER methodology was successfully introduced in this study, and two series of equations with good predictive power and stability were obtained for the prediction of aqueous solubility and *n*-octanol/water partition coefficient. It proved the applicability of TLSER method in the QSPR study of macromolecular like PCBs, PCDDs and PCDFs.

Besides their high accuracy and significant relationship between molecular structure and the predicted properties, the established models have explicit physical meanings as well as all invariables in equations.

The equation developed from the total set containing PCBs, PCDDs and PCDFs have good performance similar to or even better than equations based on individual subset. These equations can be used in the estimation of aqueous solubility and *n*-octanol/water partition coefficient without observed values yet, which also provide a promising approach for estimating the properties of other persistent organic pollutants without reported data in the current.

References:

Carbo D R, Robert D, Amat L *et al*., 2000. Molecular quantum similarity in QSAR and drug design[M]. New York: Springer-Verlag Berlin Heidelberg.

Chen J W, Quan X, Zhao Y Z *et al*., 2001. Quantitative structure-property relationship studies on *n*-octanol/water partitioning coefficients of PCDD/Fs [J]. Chemosphere, 44: 1369—1374.

Famini G R, Penski C A, Wilson L Y, 1992. Using theoretical descriptors in quantitative structure-activity relationships: some physicochemical properties [J]. J Phys Org Chem, 5: 395—406.

Govers H A J, de Voogt P, 1995a. Gas chromatographic derivation of the solubility of polychlorinated biphenyls with the inclusion of cis-trans and optical isomerism and orientational disorder[J]. SAR and QSAR in Environ Res, 3: 315—324.

Govers H A J, van der Wielen F W M, Olie K, 1995b. Derivation of solubility parameters of chlorinated dibenzofurans and dibenzo[*p*]dioxins from gas chromatographic retention parameters via SOFA[J]. J Chromatogr A, 715: 167—178.

Govers H A J, Krop H B, 1998. Partition constants of polychlorinated dibenzofurans and dibenzo-*p*-dioxins[J]. Chemosphere, 37: 2139—2152.

Karelson M, 2000. Molecular descriptors in QSAR/QSPR[M]. New York: John Wiley & Sons, Inc.

Kavlock R J, Daston G P, DeRosa C *et al*., 1996. Research needs for the risk assessment of health and environmental effects of endocrine disruptors: a report of the U. S. EPA-sponsored workshop[J]. Environmental Health Perspective, 104(supplement 4): 1—26.

Mackay D, Shiu W Y, Ma K C, 1997. Illustrated handbook of physical-chemical properties and environmental fate for organic chemicals, vols. I-V[M]. London: Boca Raton, Ann Arbor.

MINITAB Inc. 2002. MINITAB tutorials for MINITAB 13[EB]. http://www.minitab.com/resources/tutorial/index.htm

Rulle P, 2000. The *n*-octane/water partition coefficient of environmentally relevant chemicals predicted from the mobile order and disorder (MOD) thermodynamics[J]. Chemosphere, 40: 457—512.

Rulle P, Kesselring U W, 1997. Aqueous solubility prediction of environmentally important chemicals from the mobile order thermodynamics[J]. Chemosphere, 34: 275—298.

Stewart J J P, 1990. Manual of MOPAC 6.0[Z]. Frank J. Seiler Research Laboratory, US Air Force Academy, Co 80840.

Wang Y H, Wong P K, 2002. Mathematical relationships between vapor pressure, water solubility, Henry's law constant, *n*-octanol/water partition coefficient and gas chromatographic retention index of polychlorinated-dibenzo-dioxins [J]. Water Research, 36: 350—355.

Wilson L Y, Famini G R, 1991. Using theoretical descriptors in quantitative structure-activity relationships: some toxicological indices[J]. J Med Chem, 34: 1668—1674.

UNEP, 2001. Final act of the conference of plenipotentiaries on the Stockholm convention on persistent organic pollutants[C]. Stockholm, May 22—23.

Van Emon J M, Gerlach C L, Bowman K L, 2000. Analytical challenges of environmental endocrine disruptor monitoring[M]. ACS symposium series 747: Analysis of environmental endocrine disruptors (Keith L. H., Jones-Lepp T. L., Needham L. L. ed.). Washington DC: Americal Chemical Society. 23—32.

(Received for review September 9, 2002. Accepted March 10, 2003)