Bull. Environ. Contam. Toxicol. (1998) 61:591-599
© 1998 Springer-Verlag New York Inc.

Bulletin of Environmental Contamination and Toxicology

Prediction of Water Solubility and Toxicity of Substituted Indoles to *Photobacterium phosphoreum* Using Molecular Connectivity Indices and Quantum Chemical Parameters

J. Dai,[1] B. Sun,[1] A. Zhang,[1] K. Lin,[2] L. Wang[1]

[1] State Key Laboratory of Pollution Control and Resource Reuse, Department of Environmental Science and Engineering, Nanjing University, Nanjing, 210093, People's Republic of China
[2] Department of Chemistry, Nanjing University, Nanjing 210093, People's Republic of China

Received: 19 July 1998/Accepted: 13 October 1998

Indole and its derivatives are used extensively as intermediates to synthesize medicines, dyes, and pesticides. For example, indoline is used to synthesize analgesics and antihypertensives, 3-(2-aminoethyl)-indole for the synthesis of the vasodilator and antihypertensive. 2-Methylindole is a starting material for commercially important cationic diazo dyes. Cyanine dyes are synthesized from 2-methylindole, 1-methyl-2-phenylindole. In addition, indolyl-methane dyes are produced from 1-methyl-2-phenylindole. Fungicidal and bactericidal plant protectives are synthesized from indoline (Elvers *et al.* 1989). Indole and its derivatives are being introduced into the environment although behavioral and ecological effects of them are unknown.

A chemical's water solubility is one of its most important and fundamental properties in toxicology, Water solubility is critical to understanding the tendency of a chemical to cross biological membranes, In exposure assessment water solubility plays a vital role in determining the tendency of a chemical to partition from one phase to another. Thus, the ability to predict a chemical's behavior in biological or environmental systems is largely dependent upon the knowledge of its water solubility and other properties (Meylan 1996).

Luminescent bacteria toxicity test (Bulich 1986) has been widely used because it is sensitive, cost effective, and relatively simple to perform. The toxicity parameter is the EC_{50}, the concentration of toxicant that diminishes light emission by 50% at 20 °C at a given time. These EC_{50} values have been shown to correlate to bioassays utilizing guppies (*Lebistes reticulatus*), rainbow trout (*Salmo gairdneri*), fathead minnows (*Pimephales promelas*), *Spirillum,* and *Daphnia* (Chen and Que Hee 1995).

Quantitative structure-activity relationships (QSAR) are used increasingly to screen and predict toxicity and fate of chemicals released into the environment. The basic assumption of QSAR is that there is a quantitative connection between the molecular structure of compounds and their biological, chemical, and physical properties. The QSAR models are usually developed for a group of compounds

---

Correspondence to: J. Dai

with similar structure using two of the basic approaches including: (a) molecular connectivity and (b) quantum chemical descriptors. Molecular connectivity is a method describing molecular structure based solely on bonding and branching patterns rather than physical or chemical characteristics, For a given molecular structure, several types and orders of molecular connectivity indices (MCIs) can be calculated. Information of the molecular size, branching, cyclization, unsaturation, and heteroatom content of a molecule is encoded in these various indices (Kier and Hall 1976). A large number of studies have demonstrated that many physicochemical and biological properties, such as water solubility, n-octanol/water partition coefficient (Kow), soil sorption coefficient (Koc), and biological activity correlate with the connectivity index (Blum and Speece 1990; Hong *et al.* 1995). Quantum chemical descriptor is a kind of non-empirical structural descriptor which can be obtained by the chemical structure and quantum chemical software. The obvious advantages of quantum structural descriptors are: they are not restricted to closely related compounds, they can be easily obtained, and describe clearly defined molecular properties. Because of their advantages, they are widely used in the QSAR studies (Nevalainen 1994; Mekenyan *et al.* 1994).

The water solubility (Sw) for substituted indoles was determined by generator column. Toxicities of 18 substituted indoles to *Photobacterium phosphoreum* were observed. QSAR correlations using MCIs descriptors and semi-empirical quantum chemical descriptors were established. Effectiveness of regression equations in predicting Sw and toxicity to *Photobacterium phosphoreum* for the studied chemicals was compared.

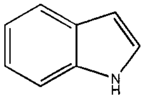
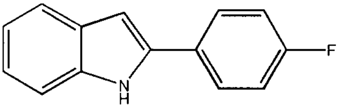
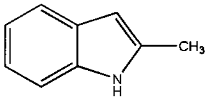
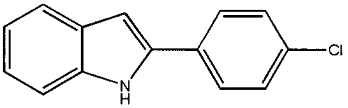
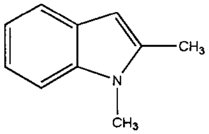
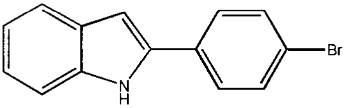
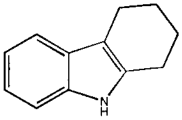
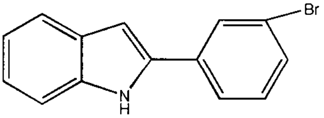
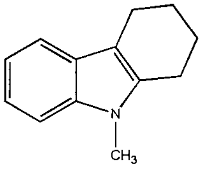
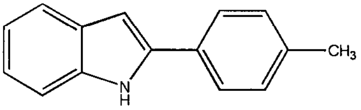
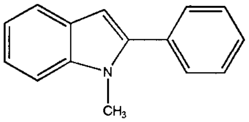
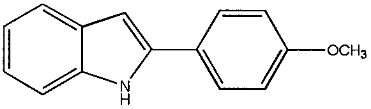
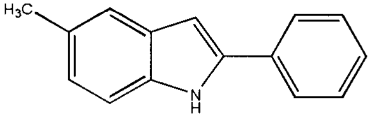
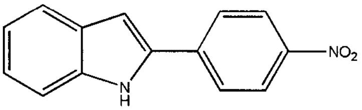
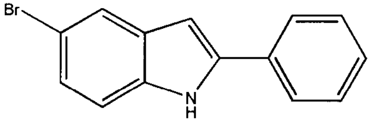
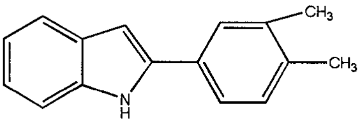
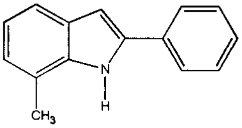
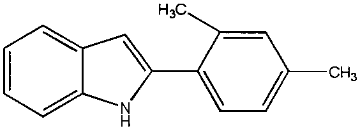
MATERIALS AND METHODS

Indole, 2-methylindole, 1,2-dimethylindole, 1,2,3,4-tetrahydrocarbazole were analytical reagent grade (A.R.), 14 substituted indoles were synthesized in our laboratory. Purities were monitored by HPLC to assure that no interference peak had occurred (Lin 1994; Table 1).

The solubilities of 18 chemicals were determined by the generator-column method described by Miller (1984) at 25±0.5°C The equilibrium concentrations for samples in water were determined by UV/visible spectrophotometer at their maximum absorption wavelengths which were obtained by scanning on Shimadzo-2201 UV/visible spectrophotometer).

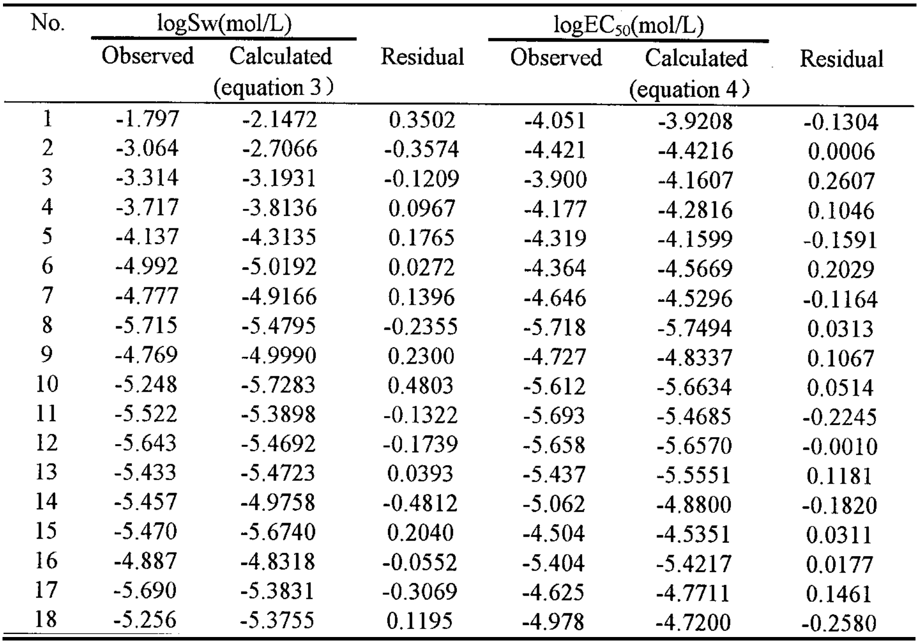
Photobacterium phosphoreum (T3 mutation) was supplied in the form of freeze dried powder by the Institute of Soil Science, the Chinese Academy of Sciences, Nanjing, P. R. China. The concentration for a 50% inhibition of bioluminescence of *P. phosphoreum* after 15 min exposure, expressed as 15min-EC_{50} (mol/L), was measured. The Microtox test instrument (Model toxicity analyzer DXY-2) was made by the Institute mentioned above. The experiment was performed at 20 °C according to the procedures described by the Instrument Manual. The test chemical

Table 1. The structures and names for the indole series of compounds

No.	Structures and names	No.	Structures and names
1	 indole	10	 2-(4’-fluoro-phenyl)-indole
2	 2-methylindole	11	 2-(4’-chloro-phenyl)-indole
3	 1,2-dimethylindole	12	 2-(4’-bromo-phenyl)-indole
4	 1,2,3,4-tetrahydrocarbazole	13	 2-(3’-bromo-phenyl)-indole
5	 N-methyl-1,2,3,4-tetrahydrocarbazole	14	 2-(4’-methyl-phenyl)-indole
6	 1-methyl-2-phenyl-indole	15	 2-(4’-methoxyl-phenyl)-indole
7	 5-methyl-2-phenyl-indole	16	 2-(4’-nitro-phenyl)-indole
8	 5-bromo-2-phenyl-indole	17	 2-(3’,4’-dimethyl)-indole
9	 7-methyl-2-phenyl-indole	18	 2-(2’,4’-dimethyl-phenyl)-indole

chemical dissolved in acetone was directly injected into the medium using a syringe injector at different volume. Solvent concentrations did not exceed recommended levels (0.1mL/L) and a solvent control equivalent to the amount the highest test concentration was used. All assays were conducted in triplicate at each concentration. The change in the intensity of bacterial luminescence in the absence (controls) and in the presence (samples) of the tested substance after incubated 15 min was recorded. In order to correct from natural variations of luminescence, each test tube (samples) was arranged between blank (controls). The 15-min percent inhibition for each concentration was calculated relative to the control and the 15-min EC_{50} was calculated by plotting probit of percent inhibition against the logarithm of concentration. (Table 2).

Table 2. Comparison of experimental data for $\log S_w$ and $\log EC_{50}$ inhibition of *P. phosphoreum* with values calculated from a correlation equation



No.	logSw(mol/L) Observed	logSw(mol/L) Calculated (equation 3)	Residual	$\log EC_{50}$(mol/L) Observed	$\log EC_{50}$(mol/L) Calculated (equation 4)	Residual
1	-1.797	-2.1472	0.3502	-4.051	-3.9208	-0.1304
2	-3.064	-2.7066	-0.3574	-4.421	-4.4216	0.0006
3	-3.314	-3.1931	-0.1209	-3.900	-4.1607	0.2607
4	-3.717	-3.8136	0.0967	-4.177	-4.2816	0.1046
5	-4.137	-4.3135	0.1765	-4.319	-4.1599	-0.1591
6	-4.992	-5.0192	0.0272	-4.364	-4.5669	0.2029
7	-4.777	-4.9166	0.1396	-4.646	-4.5296	-0.1164
8	-5.715	-5.4795	-0.2355	-5.718	-5.7494	0.0313
9	-4.769	-4.9990	0.2300	-4.727	-4.8337	0.1067
10	-5.248	-5.7283	0.4803	-5.612	-5.6634	0.0514
11	-5.522	-5.3898	-0.1322	-5.693	-5.4685	-0.2245
12	-5.643	-5.4692	-0.1739	-5.658	-5.6570	-0.0010
13	-5.433	-5.4723	0.0393	-5.437	-5.5551	0.1181
14	-5.457	-4.9758	-0.4812	-5.062	-4.8800	-0.1820
15	-5.470	-5.6740	0.2040	-4.504	-4.5351	0.0311
16	-4.887	-4.8318	-0.0552	-5.404	-5.4217	0.0177
17	-5.690	-5.3831	-0.3069	-4.625	-4.7711	0.1461
18	-5.256	-5.3755	0.1195	-4.978	-4.7200	-0.2580

Valence connectivity parameters were calculated using program developed by our laboratory (designed with FORTRAN language) according to the method described by Kier and Hall (1976): values for –Cl, –F, –Br, –O–,–N< , =0 were 0.78, 7, 0.26, 6, 5 and 6, respectively (Liao 1996). Values of MCIs were then calculated (Table 3).

Table 3. Selected molecular connectivity indices for the studied compounds

No.	$^{0}x_{p}^{v}$	$^{1}x_{p}^{v}$	$^{2}x_{p}^{v}$	$^{3}x_{p}^{v}$	$^{4}x_{p}^{v}$	$^{3}x_{pc}^{v}$	$^{4}x_{pc}^{v}$	$^{5}x_{pc}^{v}$	$^{6}x_{pc}^{v}$
1	4.9641	2.9880	2.0649	1.4625	0.9950	0.1555	0.3527	0.4516	0.3484
2	5.8868	3.4043	2.5396	1.6646	1.2185	0.2998	0.4850	0.6877	0.6260
3	6.8340	3.7991	2.9036	2.1916	1.4269	0.3888	0.9388	1.1711	1.0286
4	7.6378	5.0345	3.8348	3.0056	2.3583	0.3211	0.8398	1.3636	1.6022
5	8.5850	5.4289	4.2085	3.4341	2.7164	0.4160	1.1914	1.9746	2.3416
6	9.2207	5.4598	4.0134	3.0212	2.1748	0.4076	1.0010	1.4745	1.4886
7	9.1962	5.4761	4.1397	2.8610	2.0253	0.4777	0.8553	1.1237	1.2248
8	10.1573	5.9566	4.6947	3.1600	2.2548	0.6379	1.0278	1.3426	1.5135
9	9.1962	5.4821	4.0882	2.9493	2.0810	0.4457	0.9443	1.2696	1.3020
10	8.5741	4.9984	3.7772	2.6232	1.7953	0.3740	0.7484	0.9594	0.9240
11	9.3284	5.3755	4.2127	2.8746	1.9511	0.4997	0.8936	1.0739	1.0804
12	10.1573	5.7900	4.6912	3.1509	2.0592	0.6379	1.0531	1.1997	1.2523
13	10.1573	5.9566	4.6947	3.1580	2.2834	0.6379	1.0278	1.3599	1.4633
14	9.1962	5.3094	4.1363	2.8305	1.8990	0.4777	0.8682	1.0538	1.0530
15	9.6044	5.5884	3.9988	2.9310	1.9995	0.3790	0.8288	1.0495	1.0855
16	9.4599	5.5648	4.0742	2.9191	1.9916	0.4228	0.8721	1.0720	1.0995
17	10.1188	5.8927	4.5693	3.3479	2.1378	0.5997	1.2865	1.3880	1.4636
18	10.1188	5.8927	4.5886	3.2021	2.3343	0.6108	1.1310	1.5643	1.5360

$^{t}x_{p}^{v}$ represents *t*th order valence path index (t=0-4), $t_{pc}^{x\ V}$ represents *t*th order valence path-cluster index (*t*=3-6)

Table 4. Molecular structure descriptors derived from optimized geometric structures of the studied compounds using PM3 method.

No.	α [a.u.]	μ [a.u.]	$-E_{homo}$ [eV]	E_{lumo} [eV]	qH^{+} [a.c.u.]	$-q^{-}$ [a.c.u.]	HOF [kcal/mol]	-TE $\times 10^3$[eV]	-EE $\times 10^3$[eV]	CCR $\times 10^3$[eV]
1	73.074	0.7327	8.4023	0.1355	0.1371	0.2348	42.577	1.2171	5.8724	4.6553
2	82.925	0.7146	8.4053	0.0908	0.1277	0.1973	32.270	1.3669	7.0658	5.6989
3	91.383	0.8118	8.3135	0.1199	0.1280	0.1780	31.443	1.5162	8.4787	6.9625
4	140.504	0.6856	8.2136	0.1288	0.1054	0.2076	23.145	1.7845	10.998	9.2136
5	113.017	0.7484	8.1514	0.1366	0.1058	0.1849	22.279	1.9339	12.656	10.723
6	141.209	0.8112	8.2813	-0.268	0.1301	0.1486	65.912	2.1386	13.929	11.791
7	144.502	0.6451	8.2814	-0.338	0.1296	0.1749	56.599	2.1387	13.594	11.455
8	150.347	1.2242	8.5694	-0.558	0.1321	0.1634	72.713	2.3274	13.623	11.296
9	142.845	0.6595	8.3224	-0.393	0.1296	0.1540	56.827	2.1390	13.744	11.605
10	137.295	0.5440	8.5010	-0.614	0.1294	0.1608	22.217	2.4142	13.803	11.389
11	148.481	0.4580	8.4523	-0.588	0.1294	0.1606	59.026	2.2907	13.607	11.316
12	150.546	0.4789	8.5093	-0.608	0.1295	0.1660	73.490	2.3273	13.599	11.271
13	148.880	0.7516	8.5026	-0.595	0.1320	0.1652	73.494	2.3273	13.653	11.326
14	145.470	0.7124	8.3434	-0.391	0.1294	0.1599	56.283	2.1390	13.567	11.428
15	152.440	0.6942	8.2950	-0.340	0.1290	0.1871	27.648	2.4322	15.285	12.853
16	155.946	2.1195	8.8267	-1.438	0.1335	0.6015	56.884	2.7207	16.587	13.866
17	153.840	0.7532	8.3215	-0.360	0.1292	0.1609	47.731	2.2887	15.211	12.922
18	150.300	0.7311	8.3710	-0.216	0.1311	0.1628	48.940	2.2886	15.392	13.104

Molecular geometry was optimized and parameters were calculated using the semi-empirical orbital MOPAC6.0 procedures (Stewart 1990) according to methods PM3 (Stewart 1989). MOPAC was used to determine the individual structural parameters influencing water solubility and toxicity, including the average polarizablity (α), the dipole moment (μ), the energy of the highest occupied molecular orbital (E_{homo}), and the energy of the lowest unoccupied molecular orbital (E_{lumo}), the most positive net atomic charges on hydrogen (qH^+), the largest negative atomic charge on an atom (q^-), the heat of formation (HOF), the total energy (TE), electronic energy (EE), and core-core repulsion (CCR) (Table 4).

The regression analysis was performed using the “Statgraphics” program (STSC Inc 1985).

RESULTS AND DISCUSSION

In order to describe the relationships among logSw, $\log EC_{50}$, and calculated structural parameters, a multiparameter regression analysis was performed. The obtained QSAR equations are presented as follows:

MICs:

$$\log Sw = 1.38 - 0.76\,{}^{0}\chi^{v} + 0.47\,{}^{6}\chi^{v}_{pc}$$
$$n=18,\ R^2(\text{adj})=0.93,\ SE=0.29,\ F=113.7,\ p=0.0000 \qquad (1)$$

$$\log EC_{50} = -2.81 - 0.94\,{}^{2}\chi^{v} + 0.47\,{}^{5}\chi^{v}_{pc}$$
$$n=18,\ R^2(\text{adj})=0.65,\ SE=0.27,\ F=16.6,\ p=0.0000 \qquad (2)$$

PM3:

$$\log Sw = 0.271 - 3.95q^{-} + 0.003TE$$
$$n=18,\ R^2(\text{adj})=0.9387,\ SE=0.27,\ F=131,\ p=0.0000 \qquad (3)$$

$$\log EC_{50} = 19.85 + 3.43E_{homo} + 1.11E_{humo} + 27.36qH^{+} - 4.9q^{-}$$
$$n=18,\ R^2(\text{adj})=0.92,\ SE=0.17,\ F=51.3,\ p=0.0000 \qquad (4)$$

Where n is the number of chemicals, R is the correlation coefficient, F is the F-test value, p is the significance level of the whole equation.

The F and p values of equations (1) and (3) were significant with large correlation coefficients and small SE values. This indicated that Sw can be predicted by models based on valence connectivity indices and quantum chemical descriptors. Equations (2) and (4) predicted $\log EC_{50}$ based on valence connectivity, but, the correlation was weak with large SE compared with the quantum chemical method. This indicated that the quantum chemical method is more effective than valence connectivity methods for predicting toxicity to *P. phosphoreum* (Table 2). The predicted logSw and $\log EC_{50}$ were well fitted with observed logSw and $\log EC_{50}$ (Figs. 1 and 2).

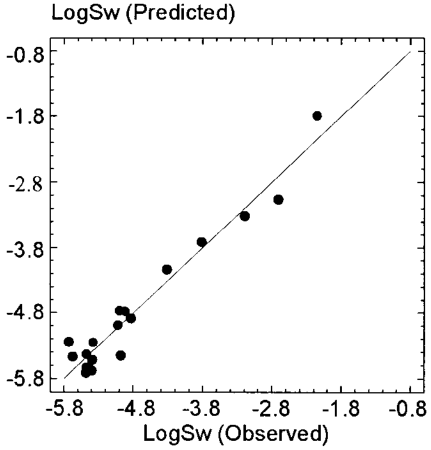


Figure 1. Plot of observed 1ogSw vs. calculated 1ogSw using equation (3)

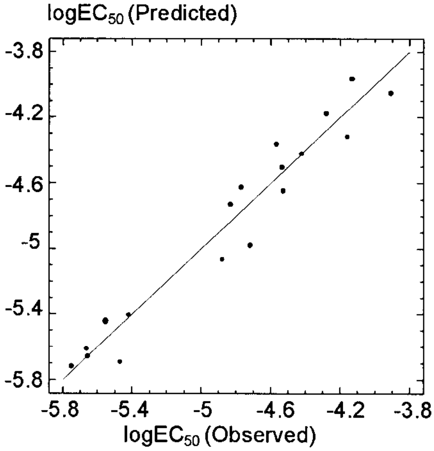


Figure 2. Plot of observed $logEC_{50}$ vs. calculated $logEC_{50}$ using equation (4)

Student t-values for partial correlation coefficients in equation 3 are 16.044, 6.007 for TE and q^-, respectively. This indicated that the total molecular energy (TE) descriptor is the most significant factor for 1ogSw. Since the TE provides information about the bulkiness of the solute, this size-related descriptor may be related to the interaction between the solute molecule and the water phase. The largest negative atomic charge on an atom (q^-) descriptor is related to the ability of the solutes to participate in hydrogen-bonding interactions with solvent molecules.

The student t-values for partial correlation coefficients in equation 4 are -8.39, 5.96, -5.74, and 4.09 for q^-, El_{umo}, Eh_{omo}, qH^+, respectively. This indicated that the largest negative atomic charge on an atom (q^-) is the most significant factor for $logEC_{50}$. In addition, E_{homo} and E_{lumo} descriptors are related to the ability of solutes to participate in electron pair donor-acceptor interactions with membrane molecules of organisms. As may be concluded from the partial correlation coefficients, this kind of interaction is less important for toxicity. The qH^+ determines the toxicity to *P. phosphoreum* for test compounds, and is related to the ability of the solutes to participate in hydrogen-bonding interactions with the membrane molecules. Its participation in the correlation are less than the Eh_{omo} and E_{lumo}, indicating that it is not important for toxicity. The parameter q^- had a negative sign, which means that the larger q^-, the smaller is the $logEC_{50}$ value. On the contrary, the parameter El_{umo}, Eh_{omo}, and qH^+ had a positive sign, which means that the larger E_{lumo}, E_{homo}, and qH^+, the larger the $logEC_{50}$ value, and the less toxicity *to P. phosphoreum.*

The results of this study indicate that the regression equations based on quantum chemical descriptors are successful in predicting 1ogSw and $logEC_{50}$ to *P.*

phosphoreum for 18 substituted indoles. The obtained equations can be used to predict of Sw and toxicity *to P. phosphoreum* for similar series of compounds.

Acknowledgments. This research was financed by the National Natural Science Fund of P. R. China.

REFERENCES

Blum DJW, Speece RE (1990) Determining chemical toxicity to aquatic species: the use of QSARs and surogate organisms. Environ Sci Technol 24: 284-293

Bulich AA (1986) Bioluminescence assays, In Toxicity Testing Using Microorganisms (Edited by G Bitton and BJ Dtka) Vol 1, CRC Press, Boca Raton, FL, pp5-74

Chen HF, Que Hee SS (1995) Ketone EC_{50} values in the Microtox test. Ecotoxicol Environ Saf 30:120-123

Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP (1985) AM1: a new general purpose quantum mechanical molecular model. J Am Chem Soc 107: 3902-3909

Elvers B, Hawkins S, Ravenscroft M, Schulz G (1989) Ullmann's Encyclopedia of Industrial Chemistry, Cambridge; New York, NY: VCH, vol. A 14, 167-170

Hong H, Han, SK, Wang XR Wang LS(1995) Prediction of partition coefficient and toxicity for phenylthio, phenyslulfinyl, and phenylsulfonyl acetates. Environ Sci Technol 29: 3044-3048

Kier LH, Hall (1976) Molecular connectivity in chemistry and drug research. Academic Press, New York

Liao YY, Wang ZT, Wang LS (1996) The prediction of soil sorption coefficients of heterocyclic nitrogen compounds by octanol/water partition coefficients, water solubility and by molecular connectivity indices. Bull Environ Contam Toxicol 56: 711-716

Ling KQ, Liu ZB (1994) Dye-sensitized photooxygenation of 2-substituted indoles. Organic Chemistry (in Chinese) 14:593-598

Mekenyan OG, Ankley GT, Veith GD, Call DJ (1994) QSARs for photoinduced toxicity: 1, Acute lethality of polycyclic aromatic hydrocarbons to Daphnia magna. Chemosphere 28:567-572

Meylan WM, Howard PH, Boethling RS (1996) Improved method for estimation water solubility form octanol/water partition coefficient. Environ Toxicol Chem 15:100-106

Miller MM, Ghodbane S, Wasol SP, Twarl YB, Martire DE (1984) Aqueous solubilities, octanol/water partition coefficients, and entropies of melting of chlorinated benzenes and biphenyls. J Chem Eng Data 29:184-190

Nevalainen T, Kolehaminen E (1994) New QSAR models for polyhalogenated aromatics. Environ Toxicol Chem 13:1699-1706

Randic M (1975) On characterization of molecular branching. J Am Chem Soc 97: 6609-6615

Stewart JJP (1990) MOPAC (Ver.6.0) Manual, Frank J seiler Research Laboratory, U.S. Air Force Academy, Co 80840

Stewart JJP (1989) Optimization of parameters for semiempirical methods I. J

Comp Chem 10:209-220

STSC Inc (1985) and Statistical Graphics Corporation, STATGRAPHICS, Version 4.0, (Serial Number: 3017435)