

Predicting Aqueous Solubility of Chlorinated Hydrocarbons by the MCI Approach

Ying-Long Wang, Yang-Dong Hu *, Lian-Ying Wu and Wei-Zhong An

College of Chemistry and Chemical Engineering, Ocean University of China, Qingdao 266003, People's Republic of China

* Author to whom correspondence should be addressed; E-mail: chem_ouc@yahoo.com.cn or ylong_wang@yahoo.com

Received: 24 November 2005 / Accepted: 17 February 2006 / Published: 28 February 2006

Abstract: Correlation for estimation of the aqueous solubility ($\log S_w$) of chlorinated hydrocarbons molecules is proposed. The MCI based quantitative structure-property relationship (QSPR) model proposed is predictive and requires only three connectivity indices in the calculation. The correlation equation obtained which is based on a training set of 50 chlorinated hydrocarbons has a correlation coefficient of 0.9670 and a standard error of 0.44 \log_{10} units. Application of the developed model to a testing set of 73 chlorinated hydrocarbons demonstrates that the new model is reliable with good predictive accuracy and simple formulation. Besides, the model does not require any experimental physicochemical properties in the calculation, so it is easy to apply, especially in cases where it is inconvenient or impossible to measure the physicochemical properties.

Keywords: molecular connectivity index; aqueous solubility; QSPR; chlorinated hydrocarbons, property model

1. Introduction

Aqueous solubility is a particularly important physicochemical property of organic chemicals that plays a significant role in various physical and biological processes, especially in drug transport and environment impact. Comparing with the time-consuming experimental procedures to determine aqueous solubility directly, reliable computational methods to predict aqueous solubility are more popular in today's research [1-3].

There are a large number of successful prediction methods, which can be divided into two main groups. The first approach [4-8] is to build model from more easily measured physicochemical properties, such as melting point, boiling point, molar volume, partition coefficient, chromatographic retention time, etc. The other method is based on the information from the molecular of the organic chemicals, which can be further divided into two classes, one is group contributions method [9-12] and the other is QSPR approach [1, 13-32].

The molecular connectivity indices which were proposed 30 years ago, have been successfully used in the correlation of various physicochemical properties of organic substances [33-35] especially in the recent applications to computational molecular design studies [36]. In the previous works, correlations of aqueous solubility using molecular connectivity indices and other descriptors have been studied and demonstrated the possibility of molecular connectivity indices in modeling aqueous solubility. In this study, we use different indices comparing with the already existing models to correlate the aqueous solubility and obtain the simpler model with the same or higher accuracy.

2. Data Set

The data set that has been studied by Eduardo J. Delgado [22] is adopted as the training set and listed in Table 2. To test the predictive ability of the proposed model, the aqueous solubility data for 73 chlorinated hydrocarbons were collected from the literature [6, 10, 27], as the testing set, and shown in Table 3. Both the training set and the testing set contain saturated, unsaturated, aliphatic and aromatic compounds, dioxins and PCBs.

3. Methods

Molecular connectivity indices have been widely used as molecular structural descriptors to correlate the physical properties of organic chemicals and used in computational molecular design studies. Recently, higher-order connectivity indices have been demonstrated the advantage of incorporating effects that are due to larger-scale structural features in a molecule on physical properties [36].

The simple and valence connectivity indices defined and developed by Randic [34, 35, 37-39], Kier [40-42] and Hall [43, 44] are used in this work which can be expressed by the following equation.

$${}^m\chi_k^p = \sum_{j=1}^{n_m} \prod_{i=1}^{m+1} (\delta_i^p)_j^{-0.5} \quad (1)$$

where m is the order of the connectivity index; k denotes a contiguous path type of fragment, which is divided into paths (P), clusters (C), etc; p denotes which type connectivity index is (simple, valence or other type); n_m is the number of the relevant paths; δ_i^p is the connectivity index.

In this work, for each chemical the values of the connectivity indices up to third order are calculated using the vertex adjacency matrix. The simple connectivity index (δ) and the valence connectivity index (δ^v) used in this study are summarized in Table 1.

Table 1. Connectivity Index (δ and δ^v) values of groups used in this work

Group	δ	δ^v	Group	δ	δ^v
-CH ₃	1	1	=CH ₂	1	2
-CH ₂ -	2	2	=CH-	2	3
>CH-	3	3	=C<	3	4
>C<	4	4	-Cl	1	7/9

The detailed equations for the simple and valence molecular connectivity indices for zeroeth, first, second, and third orders are listed as follows:

$${}^0\chi = \sum_{\text{vertices}} \frac{1}{\sqrt{\delta_i}} \quad (2)$$

$${}^0\chi^v = \sum_{\text{vertices}} \frac{1}{\sqrt{\delta_i^v}} \quad (3)$$

$${}^1\chi = \sum_{\text{edges}} \frac{1}{\sqrt{\delta_i \delta_j}} \quad (4)$$

$${}^1\chi^v = \sum_{\text{edges}} \frac{1}{\sqrt{\delta_i^v \delta_j^v}} \quad (5)$$

$${}^2\chi = \sum_{\text{triplets}} \frac{1}{\sqrt{\delta_i \delta_j \delta_k}} \quad (6)$$

$${}^2\chi^v = \sum_{\text{triplets}} \frac{1}{\sqrt{\delta_i^v \delta_j^v \delta_k^v}} \quad (7)$$

$${}^3\chi_p = \sum_{\text{quaternion-path}} \frac{1}{\sqrt{\delta_i \delta_j \delta_k \delta_l}} \quad (8)$$

$${}^3\chi_p^v = \sum_{\text{quaternion-path}} \frac{1}{\sqrt{\delta_i^v \delta_j^v \delta_k^v \delta_l^v}} \quad (9)$$

$${}^3\chi_c = \sum_{\text{quaternion-cluster}} \frac{1}{\sqrt{\delta_i \delta_j \delta_k \delta_l}} \quad (10)$$

$${}^3\chi_c^v = \sum_{\text{quaternion-cluster}} \frac{1}{\sqrt{\delta_i^v \delta_j^v \delta_k^v \delta_l^v}} \quad (11)$$

After the calculation of ten molecular connectivity indices, stepwise regression using MATLAB Statistics Toolbox [45] is used for choosing the variables and fitting the experimental data of the data set.

The average absolute error (AAE) and the root-mean-square error (RMSE) were calculated as the

following equations to compare with the existing model.

The AAE was calculated as

$$AAE = \frac{\sum |\log S_{cal} - \log S_{exp}|}{N} \quad (12)$$

The RMSE was calculated as

$$RMSE = \sqrt{\frac{\sum (\log S_{cal} - \log S_{exp})^2}{N}} \quad (13)$$

where N=number of compounds.

4. Results and Discussion

Delgado²² used CODESSA to develop QSPR model and carried out a correlation analysis to find the best QSPR model using a heuristic method. He succeeded in obtaining the two descriptors that have definite physical meaning corresponding to different intermolecular interactions.

In our work, the coefficients of the best correlation model for aqueous solubility of the 50 chlorinated hydrocarbons used as training set in this study are shown in table 2 and equation (14). The $^0\chi$ that reflects the size of the molecule is the most significant descriptor, as can be seen by its highest t-test value. This conclusion is in agreement with the existing models^{22, 27}. The other descriptors $^3\chi_c$ and $^3\chi_c^v$ that reflect the contribution of clusters in a molecule to aqueous solubility are also important in describing the aqueous solubility of chlorinated hydrocarbons. This demonstrates again that higher-order connectivity indices contain a large amount of information about the molecule, especially the larger-scale structural features (such as branching)³⁶.

Table 2. The best correlation model of logSw of 50 compounds

n	Descriptor	Coefficient	t-test
0	Intercept	0.3556	
1	$^0\chi$	-0.4373	-7.7127
2	$^3\chi_c$	-3.0857	-4.7889
3	$^3\chi_c^v$	1.9657	4.2359

The model we obtained is as the following general correlation:

$$\log Sw = 0.3556 - 0.4373^0\chi - 3.0857^3\chi_c + 1.9657^3\chi_c^v \quad (14)$$

$$R^2=0.9670, F=450.0, RMSE=0.444, n=50$$

The results calculated with equation (14) are shown in Table 3, where the experimental values and the calculated results from the Delgado method are also listed, and the scatter plot is shown in Fig 1.

The AAE for our model is 0.31 which is smaller than the 0.32 for the Delgado model, indicating that the new model has comparable accuracy to the existing model.

Table 3. Calculated results of the molar aqueous solubility of the 50 compounds (logSw).

No	CAS No.	Name	Experimental	Delgado	This work
1	75-09-2	Dichloromethane	-0.74	-0.77	-0.83
2	67-66-3	Trichloromethane	-1.19	-1.32	-1.34
3	56-23-5	Tetrachloromethane	-2.26	-1.20	-2.05
4	79-34-5	1,1,2,2-Tetrachloroethane	-1.76	-2.42	-2.27
5	540-59-0	1,2-Dichloroethene	-1.07	-1.54	-1.14
6	79-01-6	Trichloroethene	-2.04	-2.05	-2.05
7	127-18-4	Tetrachloroethene	-2.57	-2.51	-2.69
		Benzene			
8	108-90-7	Monochloro	-2.42	-2.46	-2.40
9	541-73-1	1,3-Dichloro	-3.04	-3.31	-3.30
10	95-50-1	1,2-Dichloro	-3.02	-3.17	-3.07
11	106-46-7	1,4-Dichloro	-3.31	-3.33	-3.30
12	120-82-1	1,2,4-Trichloro	-3.64	-4.11	-3.97
13	87-61-6	1,2,3-Trichloro	-4.08	-3.94	-3.77
14	108-70-3	1,3,5-Trichloro	-4.55	-4.26	-4.20
15	634-66-2	1,2,3,4-Tetrachloro	-4.38	-4.72	-4.46
16	95-94-3	1,2,4,5-Tetrachloro	-5.19	-4.94	-4.64
17	634-90-2	1,2,3,5-Tetrachloro	-4.73	-4.94	-4.67
18	608-93-5	Pentachloro	-5.37	-5.56	-5.16
		Dibenzo-p-dioxin			
19	39227-53-7	1-Chloro	-5.72	-5.66	-6.05
20	39227-54-8	2-Chloro	-5.86	-5.92	-6.24
21	29446-15-9	2,3-Dichloro	-7.23	-6.83	-6.91
22	33857-26-0	2,7-Dichloro	-7.83	-7.02	-7.14
23	39227-58-2	1,2,4-Trichloro	-7.53	-7.50	-8.11
24	30746-58-8	1,2,3,4-Tetrachloro	-8.77	-8.08	-8.63
		Biphenyl			
25	2051-60-7	2-Chloro	-4.63	-4.55	-4.66
26	2051-61-8	3-Chloro	-4.88	-5.38	-4.84
27	2051-62-9	4-Chloro	-5.25	-5.35	-4.84
28	2050-68-2	4,4'-Dichloro	-6.63	-6.05	-5.74
29	34883-39-1	2,5-Dichloro	-5.27	-5.52	-5.56
30	33284-50-3	2,4-Dichloro	-5.29	-6.28	-5.56
31	33146-45-1	2,6-Dichloro	-5.07	-4.86	-5.38
32	2050-68-2	2,4'-Dichloro	-5.60	-5.57	-5.56
33	13029-08-8	2,2'-Dichloro	-5.36	-4.76	-5.37
34	37680-65-2	2,2',5-Trichloro	-5.65	-5.60	-6.27
35	35693-92-6	2,4,6-Trichloro	-6.07	-5.93	-6.28

(Table 3 continued)

36	15862-07-4	2,4,5-Trichloro	-6.27	-6.47	-6.23
37	32598-13-3	3,3',4,4'-Tetrachloro	-8.68	-7.90	-7.09
38	35693-99-3	2,2',5,5'-Tetrachloro	-6.44	-6.33	-7.17
39	33284-53-6	2,3,4,5-Tetrachloro	-7.26	-7.04	-6.74
40	18259-05-7	2,3,4,5,6-Pentachloro	-7.78	-8.40	-7.29
41	37680-73-2	2,2',4,5,5'-Pentachloro	-7.44	-7.53	-7.84
42	55312-69-1	2,2',3,4,5-Pentachloro	-7.10	-7.21	-7.46
43	74472-44-9	2,3,3',4',5,6-Hexachloro	-7.83	-8.40	-8.35
44	55215-18-4	2,2',3,3',4,5-Hexachloro	-8.04	-9.30	-8.15
45	33979-03-2	2,2',4,4',6,6'-Hexachloro	-8.48	-8.26	-8.63
46	35065-27-1	2,2',4,4',5,5'-Hexachloro	-8.57	-8.61	-8.51
47	38411-22-2	2,2',3,3',6,6'-Hexachloro	-7.86	-7.27	-8.22
48	38380-07-3	2,2',3,3',4,4'-Hexachloro	-9.00	-8.53	-8.15
49	2136-99-4	2,2',3,3',5,5',6,6'-Octachloro	-9.30	-9.45	-9.61
50	40186-72-9	2,2',3,3',4,4',5,5',6-Nonachloro	-9.93	-9.88	-10.09

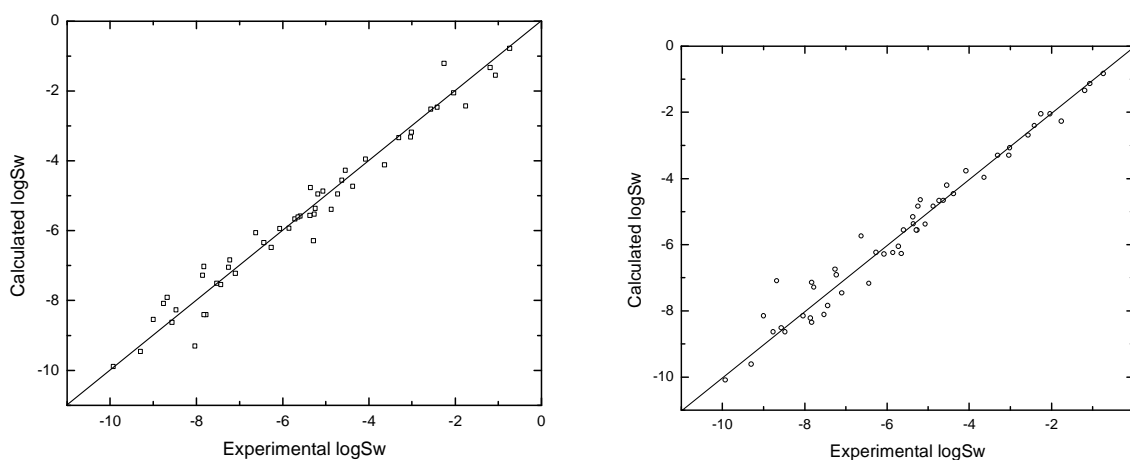


Figure 1. Calculated values versus experimental values of logSw for the two models (□: Delgado, ○: This work).

To test the predictive ability of our model, the aqueous solubility data for 73 chlorinated hydrocarbons were collected from the literature [6, 10, 27] as the testing set. The predictive results calculated with equation (14) are shown in Table 4, where the experimental values and the residual values are also listed, and the scatter plot is shown in Fig 2.

The AAE for the testing set is 0.38 and demonstrates that the proposed model is reliable and has good predictive ability.

Table 4. Predicted Results of the Molar Aqueous Solubility for 73 compounds (logSw)

No.	CAS No.	Name	Experi- mental	Calcula- ted	Residual
1	58-89-9	Lindane	-4.59	-4.77	0.18
2	16606-02-3	2,4',5-PCB	-6.25	-6.46	0.21
3	31508-00-6	2,3',4,4',5-Pentachlorobiphenyl	-7.39	-7.80	0.41
4	32598-11-1	2,3',4',5-Tetrachlorobiphenyl	-7.25	-7.13	-0.12
5	35065-28-2	2,2',3,4,4',5'-Hexachlorobiphenyl	-8.32	-8.33	0.01
6	35694-08-7	2,2',3,3',4,4',5,5'-octachlorobiphenyl	-9.16	-9.54	0.38
7	38380-02-8	2,2',3,4,5'-Pentachlorodiphenyl	-7.91	-7.66	-0.25
8	38380-08-4	2,3,3',4,4',5-Hexachlorobiphenyl	-7.82	-8.32	0.50
9	38444-85-8	2,3,4'-Trichlorobiphenyl	-6.26	-6.25	-0.01
10	41464-39-5	2,2',3,5'-Tetrachlorobiphenyl	-6.47	-6.97	0.50
11	52663-63-5	2,2',3,5,5',6-Hexachlorobiphenyl	-7.42	-8.39	0.97
12	52663-69-1	2,2',3,4,4',5',6-Heptachlorobiphenyl	-7.92	-9.06	1.14
13	52663-77-1	2,2',3,3',4,5,5',6,6'-Nonachlorobiphenyl	-10.40	-10.12	-0.28
14	52712-04-6	2,2',3,4,5,5'-Hexachlorobiphenyl	-7.68	-8.36	0.68
15	52712-05-7	2,2',3,4,5,5',6-Heptachlorobiphenyl	-8.94	-8.90	-0.04
16	55215-17-3	2,2',3,4,6-Pentachlorobiphenyl	-7.43	-7.49	0.06
17	55702-45-9	2,3,6-Trichlorobiphenyl	-6.29	-6.08	-0.21
18	56558-16-8	2,2',4,6,6'-Petachlorobiphenyl	-7.32	-7.73	0.41
19	74472-42-7	2,3,3',4,4',6-Hexachlorobiphenyl	-7.66	-8.35	0.69
20	75-09-2	dichloromethane	-0.63	-0.83	0.20
21	67-66-3	trichloromethane	-1.17	-1.34	0.17
22	56-23-5	tetrachloromethane	-2.31	-2.05	-0.26
23	75-34-3	1,1-dichloroethane	-1.29	-1.53	0.24
24	107-06-2	1,2-dichloroethane	-1.06	-1.14	0.08
25	71-55-6	1,1,1-trichloroethane	-2.00	-2.56	0.56
26	79-00-5	1,1,2-trichloroethane	-1.48	-1.75	0.27
27	79-34-5	1,1,2,2-tetrachloroethane	-1.74	-2.27	0.53
28	630-20-6	1,1,1,2-tetrachloroethane	-2.18	-2.62	0.44
29	76-01-7	pentachloroethane	-2.60	-3.06	0.46
30	2050-67-1	3,3'-PCB	-5.80	-5.74	-0.06
31	38444-81-4	2,3',5-Trichlorobiphenyl	-6.01	-6.46	0.45
32	7012-37-5	2,4,4'-PCB	-6.21	-6.46	0.25
33	38444-86-9	2',3,4-Trichlorodiphenyl	-6.29	-6.23	-0.06
34	2437-79-8	2,2',4,4'-Tetrachlorobiphenyl	-6.51	-7.17	0.66
35	41464-41-9	2,2',5,6'-Tetrachlorobiphenyl	-6.80	-7.00	0.20
36	52663-62-4	2,2',3,3',4-Pentachlorobiphenyl	-7.05	-7.46	0.41
37	52663-71-5	2,2',3,3',4,4',6-Heptachlorobiphenyl	-8.30	-8.88	0.58

(Table 4 continued)

38	2051-24-3	2,2',3,3',4,4',5,5',6,6'-Decachloro-1,1'-biphenyl	-11.62	-11.27	-0.35
39	67-72-1	hexachloroethane	-3.67	-3.76	0.09
40	540-54-5	1-chloropropane	-1.47	-1.14	-0.33
41	75-29-6	2-chloropropane	-1.41	-1.70	0.29
42	142-28-9	1,3-dichloropropane	-1.62	-1.45	-0.17
43	78-87-5	1,2-dichloropropane	-1.60	-1.87	0.27
44	96-18-4	1,2,3-trichloropropane	-1.92	-2.07	0.15
45	109-69-3	1-chlorobutane	-2.03	-1.45	-0.58
46	78-86-4	2-chlorobutane	-1.96	-1.87	-0.09
47	513-36-0	1-chloro-2-methylpropane	-2.00	-1.98	-0.02
48	541-33-3	1,1-dichlorobutane	-2.40	-2.06	-0.34
49	7581-97-7	2,3-dichlorobutane	-2.70	-2.47	-0.23
50	543-59-9	1-chloropentane	-2.73	-1.76	-0.97
51	625-29-6	2-chloropentane	-2.63	-2.18	-0.45
52	616-20-6	3-chloropentane	-2.63	-2.07	-0.56
53	594-36-5	2-chloro-2-methylbutane	-2.51	-3.35	0.84
54	544-10-5	1-chlorohexane	-3.12	-2.07	-1.05
55	319-86-8	δ -hexachlorocyclohexane	-4.51	-4.77	0.26
56	75-35-4	1,1-dichloroethylene	-1.64	-2.10	0.46
57	156-59-2	1,2-dichloroethylene	-1.30	-1.14	-0.16
58	107-05-1	3-chloropropylene	-1.36	-1.14	-0.22
59	87-68-3	hexachloro-1,3-butadiene	-4.92	-4.70	-0.22
60	77-47-4	hexachlorocyclopentadiene	-5.18	-5.59	0.41
61	95-49-8	2-chlorotoluene	-3.52	-3.11	-0.41
62	100-44-7	alpha-chlorotoluene	-2.39	-2.59	0.20
63	106-43-4	p-chlorotoluene	-3.08	-3.34	0.26
64	38444-93-8	2,2',3,3'-PCB	-7.28	-6.76	-0.52
65	32598-10-0	2,3',4,4'-PCB	-7.80	-7.13	-0.67
66	41464-40-8	2,2',4,5'-PCB	-6.57	-7.17	0.60
67	15968-05-5	2,2',6,6'-PCB	-8.03	-6.83	-1.20
68	52704-70-8	2,2',3,3',5,6-PCB	-8.60	-8.19	-0.41
69	72-54-8	DDD	-7.20	-7.44	0.24
70	50-29-3	DDT	-7.15	-8.22	1.07
71	72-55-9	DDE	-6.90	-7.74	0.84
72	91-58-7	2-chloronaphthalene	-4.14	-4.22	0.08
73	90-13-1	1-chloronaphthalene	-3.93	-4.04	0.11

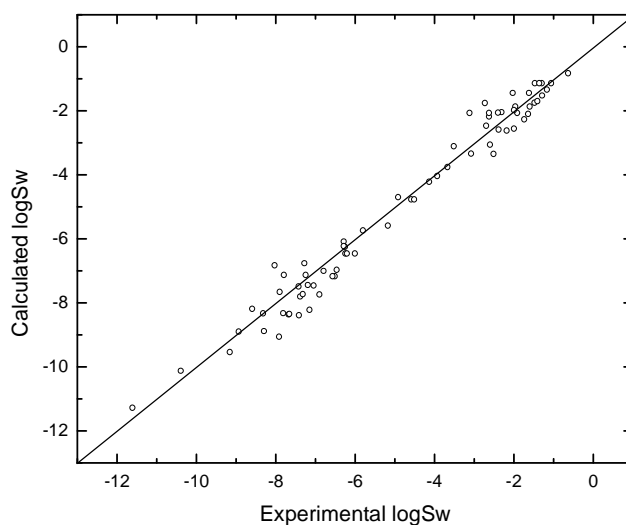


Figure 2. Calculated values versus experimental values of logSw for the testing data set.

5. Conclusion

Predictive QSPR model which is based on molecular connectivity indices is proposed in this work to correlate the aqueous solubility of 50 chlorinated hydrocarbons. Application of the developed model to a testing set of 73 chlorinated hydrocarbons demonstrates that the new model is reliable with good predictive accuracy and simple formulation. Besides, the new model does not require any experimental physicochemical properties in the calculation, so it is easy to apply, especially in cases where it is inconvenient or impossible to measure the physicochemical properties.

Acknowledgements

The authors thank the support of the Natural Science Foundation of China (Contract: 20376078) and thank the anonymous reviewers for their valuable critiques and suggestions.

References and Notes

1. Butina, D.; Gola, J. M. R. Modeling Aqueous Solubility. *J. Chem. Inf. Model.* **2003**, *43*, 837-841.
2. Erös, D.; Kéria, G.; Kövesdi, I.; Szántai-Kis, C.; Mészáros, G.; Örfi, L. Comparison of Predictive Ability of Water Solubility QSPR Models Generated by MLR, PLS and ANN Methods. *Mini-Rev. Med. Chem.* **2004**, *4*, 167-177.
3. Delaney, J. S. Predicting aqueous solubility from structure. *Drug. Discov. Today* **2005**, *10*, 289-295.
4. Yalkowsky, S. H.; Pinal, R. Estimation of the aqueous solubility of complex organic molecules. *Chemosphere.* **1993**, *26*, 1239-1261.
5. Jain, N.; Yalkowsky, S. H. Estimation of the aqueous solubility I: Application to organic nonelectrolytes. *J. Pharm. Sci.* **2001**, *90*, 234-252.

6. Peterson, D. L.; Yalkowsky, S. H. Comparison of Two Methods for Predicting Aqueous Solubility. *J. Chem. Inf. Model.* **2001**, *41*, 1531-1534.
7. Tolls, J.; Dijk, J. V.; Verbruggen, E. J. M.; Hermens, J. L. M.; Loeprecht, B.; Schuurmann, G. Aqueous Solubility-Molecular Size Relationships: A Mechanistic Case Study Using C10- to C19-Alkanes. *J. Phys. Chem. A* **2002**, *106*, 2760-2765.
8. Yang, G.; Ran, Y.; Yalkowsky, S. H. Prediction of the aqueous solubility: Comparison of the general solubility equation and the method using an amended solvation energy relationship. *J. Pharm. Sci.* **2002**, *91*, 517-533.
9. Klopman, G.; Wang, S.; Balthasar, D. M. Estimation of Aqueous Solubility of Organic Molecules by the Group Contribution Approach. Application to the Study of Biodegradation. *J. Chem. Inf. Model.* **1992**, *32*, 474-482.
10. Kühne, R.; Ebert, R.-U.; Kleint, F.; Schmidt, G.; Schüürmann, G. Group Contribution Methods to Estimate Water Solubility of Organic Chemicals. *Chemosphere.* **1995**, *30*, 2061-2077.
11. Klopman, G.; Zhu, H. Estimation of the Aqueous Solubility of Organic Molecules by the Group Contribution Approach. *J. Chem. Inf. Model.* **2001**, *41*, 439-445.
12. Hou, T. J.; Xia, K.; Zhang, W.; Xu, X. J. ADME Evaluation in Drug Discovery. 4. Prediction of Aqueous Solubility Based on Atom Contribution Approach. *J. Chem. Inf. Model.* **2004**, *44*, 266-275.
13. Nirmalakhandan, N. N.; Speece, R. E. Prediction of aqueous solubility of organic chemicals based on molecular structure. *Environ. Sci. Technol.* **1988**, *22*, 328-338.
14. Nelson, T. M.; Jurs, P. C. Prediction of Aqueous Solubility of Organic Compounds. *J. Chem. Inf. Model.* **1994**, *34*, 601-609.
15. Huibers, P. D. T.; Katritzky, A. R. Correlation of the Aqueous Solubility of Hydrocarbons and Halogenated Hydrocarbons with Molecular Structure. *J. Chem. Inf. Model.* **1998**, *38*, 283-292.
16. Huuskonen, J.; Salo, M.; Taskinen, J. Aqueous solubility prediction of drugs based on molecular topology and neural network modeling. *J. Chem. Inf. Model.* **1998**, *38*, 450-456.
17. Mitchell, B. E.; Jurs, P. C. Prediction of Aqueous Solubility of Organic Compounds from Molecular Structure. *J. Chem. Inf. Model.* **1998**, *38*, 489-496.
18. Makino, M. Prediction of aqueous solubility coefficients of polychlorinated biphenyls by use of computer-calculated molecular properties. *Environ. Int.* **1998**, *24*, 653-663.
19. Huuskonen, J. Estimation of Aqueous Solubility for a Diverse Set of Organic Compounds Based on Molecular Topology. *J. Chem. Inf. Model.* **2000**, *40*, 773-777.
20. Yaffe, D.; Cohen, Y.; Espinosa, G.; Arenas, A.; Giralt, F. A Fuzzy ARTMAP Based on Quantitative Structure-Property Relationships (QSPRs) for Predicting Aqueous Solubility of Organic Compounds. *J. Chem. Inf. Model.* **2001**, *41*, 1177-1207.
21. Liu, R.; So, S. S. Development of Quantitative Structure-Property Relationship Models for Early ADME Evaluation in Drug Discovery. 1. Aqueous Solubility. *J. Chem. Inf. Model.* **2001**, *41*, 1633-1639.
22. Delgado, E. J. Predicting aqueous solubility of chlorinated hydrocarbons from molecular structure. *Fluid. Phase. Equilibr.* **2002**, *199*, 101-107.

23. Hua, G.; Veerabahu, S.; Pil, L. Estimation of Aqueous Solubility of Organic Compounds with QSPR Approach. *Pharm. Res.* **2002**, *19*, 497-503.
24. Engkvist, O.; Wrede, P. High-Throughput, In Silico Prediction of Aqueous Solubility Based on One- and Two-Dimensional Descriptors. *J. Chem. Inf. Model.* **2002**, *42*, 1247-1249.
25. Chen, X.-Q.; Cho, S. J.; Li, Y.; Venkatesh, S. Prediction of aqueous solubility of organic compounds using a quantitative structure-property relationship. *J. Pharm. Sci.* **2002**, *91*, 1838-1852.
26. Yan, A.; Gasteiger, J. Prediction of Aqueous Solubility of Organic Compounds by Topological Descriptors. *Qsar. Comb. Sci* **2003**, *22*, 821-829.
27. Zhong, C.; Hu, Q. Estimation of the aqueous solubility of organic compounds using molecular connectivity indices. *J. Pharm. Sci.* **2003**, *92*, 2284-2294.
28. Delaney, J. S. ESOL: Estimating Aqueous Solubility Directly from Molecular Structure. *J. Chem. Inf. Model.* **2004**, *44*, 1000-1005.
29. Bergstrom, C. A. S.; Wassvik, C. M.; Norinder, U.; Luthman, K.; Artursson, P. Global and Local Computational Models for Aqueous Solubility Prediction of Drug-Like Molecules. *J. Chem. Inf. Model.* **2004**, *44*, 1477-1488.
30. Votano, J. R.; Parham, M.; Hall, L. H.; Kier, L. B.; Hall, L. M. Prediction of Aqueous Solubility Based on Large Datasets Using Several QSPR Models Utilizing Topological Structure Representation. *Chem. Biodivers.* **2004**, *1*, 1829-1841.
31. Yan, A.; Gasteiger, J.; Krug, M.; Anzali, S. Linear and nonlinear functions on modeling of aqueous solubility of organic compounds by two structure representation methods. *J. Comput. Aid. Mol. Des.* **2004**, *15*, 75-87.
32. Catana, C.; Gao, H.; Orrenius, C.; Stouten, P. F. W. Linear and Nonlinear Methods in Modeling the Aqueous Solubility of Organic Compounds. *J. Chem. Inf. Model.* **2005**, *45*, 170-176.
33. Bicerano, J. *Prediction of Polymer Properties*. Marcel Dekker: New York, 1996.
34. Randic, M. The connectivity index 25 years after. *J. Mol. Graph. Model.* **2001**, *20*, 19-35.
35. Randic, M.; Pompe, M.; Mills, D.; Basak, S. C. Variable connectivity index as a tool for modeling structure-property relationships. *Molecules.* **2004**, *9*, 1177-1193.
36. Camarda, K. V.; Sunderesan, P. An Optimization Approach to the Design of Value-Added Soybean Oil Products. *Ind. Eng. Chem. Res.* **2005**, *44*, 4361-4367.
37. Randic, M. On Characterization of Molecular Branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609-6615.
38. Randic, M.; Hansen, P. J.; Jurs, P. C. Search for usefule graph theoretical invariants of molecular structure. *J. Chem. Inf. Model.* **1988**, *28*, 60-68.
39. Randic, M.; Plavsic, D.; Lers, N. Variable Connectivity Index for Cycle-Containing Structures. *J. Chem. Inf. Model.* **2001**, *41*, 657-662.
40. Kier, L. B.; Hall, L. H.; Murray, W. J.; Randic, M. Molecular connectivity. I: Relationship to nonspecific local anesthesia. *J Pharm Sci* **1975**, *64*, 1971-1974.
41. Kier, L. B.; Hall, L. H. *Molecular connectivity in chemistry and drug research*. Academic Press: New York, 1976.
42. Kier, L. B.; Hall, L. H. *Molecular connectivity in structure-activity analysis*. Wiley: New York, 1986.

43. Hall, L. H.; Kier, L. B. Structure-activity studies using valence molecular connectivity. *J. Pharm. Sci.* **1977**, *66*, 642-644.
44. Hall, L. H.; Kier, L. B. Molecular connectivity and substructure analysis. *J. Pharm. Sci.* **1978**, *67*, 1743-1747.
45. Martinez, W. L.; Martinez, A. R. *Computational Statistics Handbook with MATLAB*. Chapman & Hall/CRC: Boca Raton London New York Washington, D.C., 2002.

© 2006 by MDPI (<http://www.mdpi.org>). Reproduction is permitted for noncommercial purposes.