

Predicting octanol/water partition coefficient using solvation free energy and solvent-accessible surface area

LIU Xin-hui, WU Chun-de, HAN Shuo-kui, WANG Lian-sheng

(State Key Laboratory of Pollution Control and Resource Reuse, Department of Environmental Science, School of the Environment, Nanjing University, Nanjing 210093, China. E-mail: ecl@nju.edu.cn)

Abstract: The regression model for octanol/water partition coefficients (K_{ow}), is founded with only two molecular descriptors available through quantum chemical calculations: solvation free energy (ΔG_s), and solvent-accessible surface area (SASA). For the properties of 47 organic compounds from 17 types, the model gives a correction coefficient (adjusted for degrees of freedom) of 0.959 and a standard error of 0.277 log unit. It is a suitable way to predict the partition properties that are related to solute-solvent interactions in the water phase.

Key words: solvation free energy; solvent-accessible surface area; quantum chemical descriptor

Introduction

The physicochemical profile of organic compounds largely determines their distribution between environmental media. For the environmental behavior of organic contaminants, the octanol/water partition coefficient (K_{ow}) is more significant. K_{ow} has been widely related to biochemical and/or biological activity in a number of quantitative structure-activity relationships (QSAR) (Leo, 1971).

An extensive overview and discussion of different approaches to calculate $\log K_{ow}$ values was given by Leo (Leo, 1993). In general, it is more useful to use descriptors derived mathematically from either the two-dimensional or the three-dimensional molecular structure, since this allows any relationship so derived to be extended to the prediction of the properties for expected compounds. Considerable efforts were made to calculate K_{ow} values by applications of computational chemistry methods. Most methods are based on atomic charges and a surface area (Bodor, 1989; 1991; 1992). Other groups suggested utilizing descriptors that are in some way related to the parameters of the LSER method (Brinck, 1993; Famini, 1994; Haeberlein, 1997; Eisefeld, 1999).

Continuum solvation models have been introduced to calculate free energies and enthalpies of aqueous solvation within semiempirical schemes (Cramer, 1992a, 1992b; Klamt, 1993). As solvation free energy reflects the macroscopic interaction of a compound with the medium water, it opens a way to estimate compound properties (Schüttmann, 1995a; 1995b) that are related to solute-solvent interactions in the water phase.

In the present study, we correlate $\lg K_{ow}$ values of 47 various compounds with only two molecular descriptors available from quantum chemical calculations through simple regression analysis. The equation obtained is quite successful, bearing in mind a wide variety of organic compound classes: hydrocarbon, alcohol, aldehyde, acid, nitrile, amine, ether, ester, halide, thiol, phenol, sulfide, and nitro. Moreover, both liquids and solids (at room temperature) are represented in the set.

1 Theory

The regression model derived below for $\log K_{ow}$ is based on only two molecular descriptors calculated from chemical structure: free energy of aqueous solvation, ΔG_s , and solvent-accessible surface area, SASA. Solvation free energy (ΔG_s) is the free energy for transferring solute from the gas phase to water solution, and can be separated into three terms: an electrostatic term, a term representing cavity formation energy, and a term for solute-solvent dispersion interactions. The solvent-accessible surface area (SASA) is defined as the locus of the center of a water molecule sphere of radius 1.4\AA ($1\text{\AA} = 10^{-10}\text{m}$) as it is rolled over the van der Waals surface of the solute. The accessible surface area equals the surface area of the cavity which must be created in the solvent to accommodate the solute, and is directly related to the energy of cavity formation and solute-solvent dispersion interaction (Hermann, 1972) —a quantity which is of primary importance to partitioning of solute. The starting point of the theoretical derivation is the reasoning that K_{ow} characterizes equilibrium partitioning of a compound between different phases. The following qualitative discussion about solution processes is specially structured to enable a rationalization of the predictive capabilities of solvation free energy and surface area (Pearlman, 1986).

In the process of solute transferring from the gas phase to solution, a cavity is first created in solvent, then the solute

molecule is placed into cavity, thus forming solution. Clearly, for any solvent, the free energy for transferring solute from the gas phase to solution can be written as:

$$\Delta G_{g \rightarrow \text{soln}}^0 = \Delta G_{\text{cav}}^0 + \Delta G_{\text{si}}^0. \quad (1)$$

The free energy of cavity formation, ΔG_{cav}^0 , can be estimated as:

$$\Delta G_{\text{cav}}^0 = \sigma \cdot (\text{SASA}), \quad (2)$$

where SASA is solvent-accessible surface area and σ is an effective interfacial tension. Thus, the free energy of cavity formation depends upon the chemical nature of the solvent and the size of the solute. The free energy of solute-solvent interaction, ΔG_{si}^0 , can be written as:

$$\Delta G_{\text{si}}^0 = \sum k_i V_i, \quad (3)$$

where the k_i is the empirically determined parameter, and V_i is the volume of the i -th substituent group on the solute. Thus, Eq. (1) can become Eq. (4).

$$\Delta G_{g \rightarrow \text{soln}}^0 = \sigma \cdot (\text{SASA}) + \sum k_i V_i. \quad (4)$$

For octanol solution, the equation may be written as (Schüürmann, 1995a):

$$\Delta G_{g \rightarrow 0}^0 = \alpha \cdot (\Delta G_s^0)^\gamma + \sigma \cdot \text{SASA} - \beta \cdot \text{SASA}, \quad (5)$$

where ΔG_s is the free energy for transferring solute from the gas phase to water solution, σ , β and γ are constants to be determined empirically.

Accordingly, the octanol/water partition coefficient, K_{ow} , can be estimated with the two descriptors, ΔG_s and SASA.

2 Materials and methods

The test set of 47 organic compounds is listed in Table 1, together with octanol/water partition coefficients at atmospheric pressure and 298K, solvation free energy (ΔG_s) and solvent-accessible surface area (SASA). Experimental data of $\log K_{ow}$ were taken from the literature (Lide, 1997; Wang, 1999). ΔG_s values were taken from the literature (Cramer, 1992a; Schüürmann, 1995a); in calculating ΔG_s , the continuum solvation model AM1-SM2 (Cramer, 1992a) was used to simulate solvation of the molecules in water. SASA values were calculated using CS CHMOFFICE 4.5 software package CS Chem3D program MOPAC 97 (CambridgeSoft Corp., 1986–1997), and the optimum geometries were obtained by AM1 and COSMO method (Klamt, 1993). Multilinear regression analysis for $\log K_{ow}$ was performed through SPSS 8.0 software package (SPSS Inc., 1989–1997) with ΔG_s and SASA.

Table 1 Compounds and physicochemical properties

No. Compound	lgK _{ow}						ΔG _S	SASA
	Expt.	ClogP		Eq. (6)				
		Calc.	Dev.	Pred.	Res.			
1 Hexane	4.00	3.87	0.13	3.62	0.38	9.20	69.20	
2 Heptane	4.50	4.40	0.10	4.38	0.12	10.45	76.67	
3 2,4-Dimethylpentane	3.63	4.14	-0.51	4.01	-0.38	10.03	72.85	
4 Octane	5.15	4.93	0.22	5.11	0.04	11.29	84.08	
5 Methylcyclohexane	3.88	3.87	0.01	3.23	0.65	8.79	65.16	
6 2-Methylpropene	2.35	2.20	0.15	1.80	0.55	4.60	52.33	
7 1,3-Butadiene	1.99	1.90	0.09	1.70	0.29	2.51	52.60	
8 1-Hexyne	2.73	2.51	0.22	3.00	-0.26	0.00	68.53	
9 1-Butadiene	0.86	0.92	-0.06	1.05	-0.19	-19.65	60.42	
10 Dimethylamine	-0.38	-0.52	0.14	-0.23	-0.15	-17.97	45.23	
11 Propanenitrile	0.16	0.13	0.03	0.47	-0.31	-15.88	51.52	
12 Nitroethane	0.18	0.25	-0.07	0.40	-0.22	-18.39	52.38	
13 1-Nitropropane	0.87	0.77	0.10	1.24	-0.37	-15.88	59.95	
14 Ethanol	-0.30	-0.24	-0.06	-0.55	-0.25	-20.48	43.40	
15 1-Propanol	0.25	0.29	-0.04	0.23	0.02	-19.23	51.17	
16 Prop-2-en-1-ol	0.17	-0.19	0.36	0.09	0.08	-20.48	50.38	
17 Butanoic acid	0.79	0.82	-0.03	0.73	0.06	-26.33	61.41	
18 Propanal	0.59	0.30	0.29	0.31	0.28	-15.05	49.17	

Table 1 (Cont'd)

No.	Compound	lgK _{ow}						SASA
		Expt.	ClogP		Eq. (6)		ΔG _S	
			Calc.	Dev.	Pred.	Res.		
19	Butanal	0.88	0.83	0.05	1.08	- 0.20	- 13.79	56.82
20	Methanethiol	0.78	0.65	0.13	0.57	0.21	- 3.34	44.13
21	Dimethyl sulfide	0.92	0.84	0.08	0.96	- 0.04	- 6.69	50.69
22	Tetrafluoromethane	1.18	1.11	0.07	1.51	- 0.33	14.21	42.66
23	1,1-Difluoroethane	0.75	0.90	0.15	0.97	- 0.22	1.25	45.45
24	2,2,2-Trifluoro-ethanol	0.41	0.52	- 0.11	0.35	- 0.06	- 18.81	52.21
25	Chlorofluoromethane	0.51	0.81	- 0.30	0.75	- 0.24	- 2.93	45.91
26	E-1,2-dichloroethene	1.93	1.77	0.16	2.13	- 0.20	- 2.93	60.96
27	Chloroform	1.97	1.95	0.02	2.33	- 0.38	- 5.02	64.58
28	Bromomethane	1.19	1.08	0.11	0.93	0.26	- 2.93	47.84
29	Iodoethane	2.00	2.00	2.00	2.39	- 0.39	- 2.93	63.82
30	1-Iodopropane	2.89	2.52	0.37	2.94	- 0.05	- 1.67	69.03
31	Benzene	2.13	2.14	- 0.01	1.99	0.14	- 2.09	58.93
32	Toluene	2.73	2.64	0.09	2.66	0.07	- 1.25	65.69
33	Chlorobenzene	2.84	2.86	- 0.02	2.78	0.06	- 4.80	69.36
34	1,2-Dichlorobenzene	3.38	3.45	0.07	3.46	- 0.08	7.10	78.43
35	1,4-Dichlorobenzene	3.38	3.57	- 0.19	3.67	- 0.29	- 5.80	79.87
36	Bromobenzene	2.99	3.01	- 0.02	3.01	- 0.02	- 8.30	74.23
37	1,2-Dibromobenzene	4.07	3.67	0.40	3.83	0.24	- 14.30	87.27
38	1,4-Dibromobenzene	4.07	3.87	0.20	4.09	- 0.02	- 13.40	89.55
39	Aniline	0.90	0.91	- 0.01	1.14	- 0.24	- 24.24	64.53
40	Phenol	1.48	1.47	0.01	0.97	0.49	- 24.24	62.80
41	<i>o</i> -Cresol	1.98	1.97	0.01	1.75	0.23	- 20.50	68.63
42	<i>m</i> -Cresol	1.98	1.97	0.01	1.67	0.31	- 23.30	69.65
43	<i>p</i> -Cresol	1.97	1.97	0.00	1.69	0.28	- 23.00	69.63
44	Anisole	2.11	2.06	0.05	2.53	- 0.41	- 9.80	69.96
45	Thiophenol	2.52	2.53	- 0.01	2.34	0.18	- 13.40	70.33
46	Nitrobenzene	1.85	1.88	- 0.03	1.85	0.00	- 21.90	70.70
47	Benzaldehyde	1.48	1.50	- 0.20	1.78	- 0.30	- 19.70	68.44

Notes: ΔG_s is in kJ/mol, ΔG_s data were taken from the literatures (Cramer, 1992a; Schürmann, 1995a); solvent-accessible surface area (SASA) in Å² expt. is experimental $\lg K_{ow}$ data (Lide, 1997; Wang, 1999); pred. is predicted $\lg K_{ow}$ value; res. is residual between experimental and predicted values; calc. is calculated value using ClogP (BioByte Corp., 1995); dev. is deviation between experimental and calculated values.

3 Results and discussions

The correlation equations for all 47 compounds is obtained using multilinear regression analysis:

$$\lg K_{ow} = -3.250 + 0.061\Delta G_s + 0.091 \text{SASA}, \quad (6)$$

$$(n = 47, r_{adj}^2 = 0.959, SE = 0.277, F = 534.82, P < 0.001),$$

where n represents the number of samples, r_{adj}^2 is the corrected squared correlation coefficient for degrees of freedom, SE is the standard error of estimates for the full regression, F is the F -test value, p is the significance level of F -value. t -test for regression coefficients is given in Table 2. The results reflect that the two variables are significant at the confidence level of 95%. According to the principle of statistics, a regression equation is of no significance if the applied variables are interrelated. As it is shown by the results in Table 3, the two variables have cleared up the possibility. In addition, the plot with the predicted and observed values is shown in Fig. 1, the predicted values and residuals between the predicted and observed values are listed in Table 2.

Table 2 Model fitting results for $\lg K_{ow}$

Independent variable	Coefficient	SE	<i>t</i> -value	<i>p</i>
Constant	-3.250	0.219	-14.882	<0.001
SASA	0.061	0.004	16.779	<0.001
ΔG_s	0.091	0.003	26.914	<0.001

Table 3 Correlation coefficients matrix, covariance matrix for independent variables

Independent variable	Covariance matrix		Covariance matrix	
	SASA	ΔG_s	SASA	ΔG_s
SASA	1.000	-0.065	1.147E-05	-8.01E-07
ΔG_s	-0.065	1.000	-8.01E-07	1.334E-05

In the regression model for $\lg K_{ow}$ (Eq. 6), the positive signs of regression coefficients of descriptors confirm that increasing free energy of solvation (ΔG_s), as well as both increasing dispersion interaction and cavity formation energy (SASA), lead to increase in hydrophobicity of the compounds; otherwise, lead to decrease.

For $\lg K_{ow}$ values at a range between -0.38 and 5.15, the higher correlation coefficient of 0.959 and lower standard error of 0.277 log unit indicate a good suitability for the two quantum chemical descriptors estimating octanol/water partition coefficients. Inspection of Table 1 reveals that no significant underestimations or overestimations of $\lg K_{ow}$ are observed, residuals are between -0.41 and 0.65 log units, and only four compounds showing greater absolute deviations more than 0.40. The correlation coefficient of 0.985 between the fitted and observed values (Fig. 1) also shows that the equation is more successful. From the theoretical considerations, some error compensation by linear combination of calculated ΔG_s and SASA may be a possible nonlinearity between ΔG_s and ΔG_s , or an incomplete account of the cavity formation energy in water, or a corresponding calculation error with dispersion interactions in the aqueous phase.

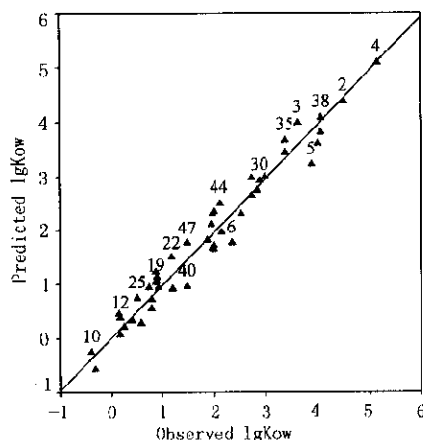
Table 1 also compares the calculated values by the widely used, empirical method, CLOGP. Since all of the compounds were included in the CLOGP basis set, the calculated values fit well with the experimental ones. However, the prediction of molecular property based on simple empirical fragment values has no scientific basis because fragments generally behave quite differently in different molecules. Therefore, the theoretical method is more reasonable and reliable.

4 Conclusions

With the development of continuum solvation models, it has become possible to link quantum chemistry to macroscopic thermodynamic properties of compounds that govern basic transfer processes of compounds between different phases. The prediction model based on mechanistic considerations need further study and evaluation with data of a large number of chemicals, to see whether additional terms will be needed for the model. The practical scope of this approach will partly depend upon the performance of the quantum chemical model used for calculating the descriptors. Therefore, application of molecular descriptors representing distinct intermolecular interactions will lead to new mechanistic insights for the structural dependence of compound properties.

References:

- Bio Byte Corp[Z], 1995.
- Bodor N, Gabanyi Z, Wong C, 1989. A new method for the estimation of partition coefficient[J]. *J Am Chem Soc*, 111: 3783—3786.
- Bodor N, Huang M J, 1991. Predicting partition coefficients for isomeric diastereoisomers of some tripeptide analogs[J]. *J Comput Chem*, 12: 1182—1186.
- Bodor N, Huang M J, 1992. An extended version of a novel method for the estimation of partition coefficients[J]. *J Pharm Sci*, 81: 272—281.
- Brinck T, Murray J S, Politzer P, 1993. Octanol/water partition coefficients expressed in terms of solute molecular surface areas and electrostatic potentials[J]. *J Org Chem*, 58: 7070—7073.
- Cambridge Soft Corp[Z], 1986—1997.
- Cramer C J, Truhlar D G, 1992a. A SCF solvation model for the hydrophobic effect and absolute free energies of solvation[J]. *Science*, 256: 213—217.
- Cramer C J, Truhlar D G, 1992b. A general parameterization for including aqueous solvation effects in the PM3 molecular orbital model[J]. *J Comput Chem*, 13: 1089—1097.
- Eisfeld W, Maurer G, 1999. Study on the correlation and prediction of octanol/water partition coefficients by quantum chemical calculations [J]. *J Phys Chem B*, 103: 5716—5729.
- Famini G R, Wilson L Y, 1994. Quantitative treatments of solute/solvent interactions[M] (Politzer P and Murray J S ed.). Amsterdam:

Fig.1 Plot of observed vs. predicted $\lg K_{ow}$

Elsevier.

- Haberlein M, Brinck T, 1997. Prediction of water-octanol partition coefficients using theoretical descriptors derived from the molecular area and the electrostatic potential[J]. *J Chem Soc Perkin Trans 2*, 289—294.
- Herrmann R B, 1972. Theory of hydrophobic bonding II: The correlation of hydrocarbon solubility with solvent cavity surface area[J]. *J Phys Chem*, 76: 2754—2759.
- Klamt A, Schürmann G, 1993. COSMO: A new approach to dielectric screening in solvents with explicit expression for the screening energy and its gradient[J]. *J Chem Soc Perkin Trans 2*, 799—805.
- Leo A, Hansch C, Elkins D, 1971. Partition coefficients and their uses[J]. *Chem Rev*, 71: 525—616.
- Leo A J, 1993. Calculating log Poct from structures[J]. *Chem Rev*, 93: 1281—1306.
- Lide D V, 1997. CRC Handbook of chemistry physics 78th[M]. Boca Raton, New York: CRC Press. 16:39—42.
- Pearlman R S, 1986. Partition coefficient determination and estimation[M] (Dunn III. W J, Block J H, Pearlman R S ed.). New York: Pergamon Press. 3—20.
- Schürmann G, 1995a. Quantum chemical approach to estimate physicochemical compound properties: application to substituted benzene[J]. *Environ Toxicol Chem*, 14: 2067—2076.
- Schürmann G, 1995b. Quantum chemical estimation of octanol/water partition coefficient—first results with aromatic phosphorothionates[J]. *Fresenius Environ Bull*, 4: 238—243.
- SPSS Inc[Z], 1989—1997.
- Wang L S, Han S K, 1999. Quantitative structure-activity relationship[M]. Beijing: Chemical Industry Press.

(Received for review June 29, 2000. Accepted September 1, 2000)