

CORRELATIONS BETWEEN OCTANOL/WATER PARTITION COEFFICIENTS AND INTRINSIC MOLECULAR PROPERTIES OF SIMPLE ORGANIC COMPOUNDS

TAKAHIRO SUZUKI*, KAZUHISA OHTAGUCHI AND KOZO KOIDE

Department of Chemical Engineering, Tokyo Institute of Technology, Tokyo 152

Key Words: Physical Property, Partition Coefficient, Hydrophobicity, Lipophilicity, Octanol, Water, Atomic Charge

Introduction

Legislation in many countries requires a manufacturer of chemicals to include physicochemical properties of molecules in its specifications in recent years. Of these properties, the partition coefficient between 1-octanol and water phases is one of the most frequently used properties. A linear relationship between the logarithms of the octanol/water partition coefficients of organic compounds (usually expressed as $\log P$) and their aqueous solubilities is reported³. In a previous paper⁹ it was shown that the aqueous solubilities of a diverse set of organic compounds can be expressed as a linear function of their molecular descriptors. The results of that study revealed two leading descriptors, the parachor as a bulk or volume term and an electronic parameter, a charge of the most negative atom in the molecule. The present work is an extension of this earlier approach to the prediction of the 1-octanol/water partition coefficients of simple organic compounds.

1. Methods

Multiple regression analysis was used, as in earlier work⁸⁻⁹ for correlating the partition coefficients with molecular descriptors. A set of 167 simple organic compounds containing carbon, hydrogen, oxygen, nitrogen, sulfur, and halogens, and representing most principal classes was considered in this study. Four molecular descriptors whose significance for correlating the aqueous solubility was found⁹ — the parachor (PA), the critical pressure (P_c), the sum of negative charges in the molecule (Q_T^-), and the charge of the most negative atom in the molecule (q_m^-), were computed as in an earlier paper⁹. The observed $\log P$ values of the 167 compounds were taken from the list of recommended values by Sangster⁷.

2. Results and Discussion

A number of different correlation equations were tested by using stepwise linear regression analysis, and a

model containing six descriptors, shown in **Table 1**, was found to be the best. The first four descriptors were also used for the aqueous solubility⁹. The descriptor, Q_T^- , was found to be statistically insignificant, so it was omitted for this case. Two additional descriptors, the number of nitrogen atoms, N_N , and the number of sulfur atoms, N_S , were newly introduced in this problem. There was no need to include such parameters for oxygen and other atoms. The relative significance of the descriptors

Table 1. Multiple linear regression model developed for 167 organic compounds ($n = 167$, $r = 0.973$, $s = 0.370$; regression range, $-1.51 \leq \log P \leq 6.90$).

descriptor	regression coefficient	F values of coeff.
PA	0.00972	352
P_c	-0.0223	20
q_m^-	5.75	667
N_F	0.969	19
N_N	-0.925	158
N_S	-0.824	43
intercept	1.72	

Table 2. Summary of regression results for the data set

Chemical class	Number of compounds	Data range	Av. absolute error*
Hydrocarbons	41	C ₅ -C ₂₄	0.32
Ethers	9	C ₄ -C ₁₂	0.38
Alcohols and Phenols	24	C ₄ -C ₁₂	0.27
Aldehydes and Ketones	11	C ₃ -C ₈	0.31
Esters	8	C ₃ -C ₉	0.20
Acids	10	C ₂ -C ₁₂	0.29
Halogenated compounds	19	C ₁ -C ₁₂	0.16
Amines	13	C ₁ -C ₁₂	0.31
Nitriles	7	C ₃ -C ₈	0.16
Nitro compounds	8	C ₁ -C ₇	0.29
Amides	7	C ₁ -C ₉	0.13
Sulfur compounds	10	C ₃ -C ₁₂	0.49
total	167	C ₁ -C ₂₄	0.29

*Av. absolute error = $\sum |\log P_{(\text{obsd})} - \log P_{(\text{estd})}| / \text{number of compounds}$.

* Received August 31, 1992. Correspondence concerning this article should be addressed to T. Suzuki.

Table 3. Subset correlations for the compound in the data set

Chemical class	Correlation	<i>n</i>	<i>r</i>	<i>s</i>
Hydrocarbons	$\log P = 0.00473PA - 0.0878P_c + 13.0q_m^- + 5.9$	41	0.970	0.30
Ethers	$\log P = 0.0166PA - 2.8$	9	0.961	0.37
Alcohols and Phenols	$\log P = 0.0236PA - 0.0306P_c + 11.7Q_T^- + 7.76q_m^- + 5.5$	24	0.984	0.23
Aldehydes and Ketones	$\log P = 0.0133PA - 79.76q_m^- - 31.0$	11	0.953	0.18
Esters	$\log P = 0.0223PA + 0.0259P_c + 6.12Q_T^- + 46.4q_m^- + 14.8$	8	0.999	0.02
Acids	$\log P = 0.0350PA + 20.8Q_T^- + 182q_m^- + 71.4$	10	0.987	0.27
Halogenated compounds	$\log P = 0.0121PA - 0.263$	19	0.967	0.19
Amines	$\log P = 0.0160PA + 0.0233P_c - 3.8$	13	0.987	0.22
Nitriles	$\log P = 0.0119PA + 58.4q_m^- + 10.5$	7	0.998	0.08
Nitro compounds	$\log P = -0.0723P_c + 130q_m^- + 21.2$	8	0.994	0.13
Amides	$\log P = 0.0120PA - 2.83$	7	0.999	0.07
Sulfur compounds	$\log P = 0.00737PA - 3.60Q_T^- + 13.1q_m^- + 0.4$	10	0.997	0.14

in the model was in the following order from their individual *F* values: $q_m^- > PA > N_N \gg N_S > P_c, N_F$. The significance of the parameter N_N , can be explained by the special role of the partitioning behavior in hydrogen bonding.

Average absolute errors for individual classes of the 167 compounds by the above model are shown in **Table 2**. Although the error for sulfur compounds is somewhat larger than that of the precision of many experimental uncertainties of $\log P$ (about 0.4 log units), the fitness of the model is satisfactory.

To examine the roles of the descriptors in more detail, the subset correlations for individual families of the data set were tested. The results are shown in **Table 3**. The significance of the two descriptors *PA* and q_m^- can also be confirmed as key parameters for $\log P$ by this result. These observations are in accord with the consensus of opinion on the partition coefficient and the solubility. It is interesting to note that *PA* is not required for nitro compounds.

From a practical viewpoint, the general correlation equation for all classes of organic compounds studied is recommended. Since the present model is not optimal for all compounds, the subset correlation equation for sulfur compounds will be useful for covering such drawbacks of the general correlation equation. The accuracy of the estimated values of the experimental descriptors *PA* and P_c , by using group contribution methods of Sugden⁵⁾ and Lydersen^{5, 6)} respectively, is satisfactory for this approach. The electronic parameters q_m^- and Q_T^- can be calculated using the RESCHA program¹⁾.

Klopman and Iroff⁴⁾ proposed a model for utilizing atomic charge densities as determined by quantum chemical calculations and structural (atom or functional group) descriptors to calculate $\log P$ values. Bodor *et al.*²⁾ presented a highly nonlinear regression model for $\log P$ by modifying Klopman's model. Compared to such approaches, there are a number of advantages to using this method: the number of descriptors is markedly fewer, and the calculation of descriptors is easier and not time-consuming since our approach employs an empirical method for calculating atomic charges in which the optimization of the molecular geometry is not required.

In addition, Klopman's and Bodor's methods give no information on applicability to compounds containing sulfur. Additional work will be required for application of the present approach to calculating $\log P$ values of multifunctional compounds.

Conclusions

A predictive equation for the 1-octanol/water partition coefficient in terms of fundamental physicochemical properties of organic compounds was developed. This method is easy to use because of its employment of calculable molecular descriptors, and it is useful for an alternative estimation scheme for $\log P$ by the group contribution approach¹⁰⁾.

Nomenclature

N_F	= number of fluorine atoms	[-]
N_N	= number of nitrogen atoms	[-]
N_S	= number of sulfur atoms	[-]
<i>n</i>	= number of data points in set	[-]
<i>PA</i>	= parachor	$[(\text{cm}^3 \cdot \text{mol}^{-1}) (\text{dyn} \cdot \text{cm}^{-1})^{1/4}]$
P_c	= critical pressure	[atm]
<i>P</i>	= 1-octanol/water partition coefficient at 25°C	[-]
Q_T^-	= sum of negative atomic charges	[electron]
q_m^-	= atomic charge of the most negative atom	[electron]
<i>r</i>	= multiple correlation coefficient	[-]
<i>s</i>	= standard deviation	[-]

Literature Cited

- 1) Baumer, L., G. Sala and G. Sello: *Tetrahedron Comput. Method.*, **2**, 37 (1989)
- 2) Bodor, N., Z. Gabanyi and C.-K. Wong: *J. Am. Chem. Soc.*, **111**, 3783 (1989)
- 3) Hansch, C., Quinlan, J.E. and Lawrence, G.L.: *J. Org. Chem.*, **33**, 347 (1968)
- 4) Klopman, G. and L.D. Iroff: *J. Comput. Chem.*, **2**, 157 (1981)
- 5) Lyman, W. J., W.F. Reehl and D.H. Rosenblatt: "Handbook of Chemical Property Estimation Methods", McGraw-Hill, New York (1982)
- 6) Reid, R.C., J.M. Prausnitz and T.K. Sherwood: "The Properties of Gases and Liquids", 3rd ed., McGraw-Hill, New York (1977)
- 7) Sangster, J.: *J. Phys. Chem. Ref. Data*, **18**, 1111 (1989)
- 8) Suzuki, T., K. Ohtaguchi and K. Koide: *J. Chem. Eng. Japan*, **25**, 434 (1992)
- 9) Suzuki, T., K. Ohtaguchi and K. Koide: *J. Chem. Eng. Japan*, **25**, 729 (1992)
- 10) Suzuki, T. and Y. Kudo: *J. Comput.-Aided Mol. Design*, **4**, 155 (1990)