Russian Journal of General Chemistry, Vol. 71, No. 4, 2001, pp. 500–504. Translated from Zhurnal Obshchei Khimii, Vol. 71, No. 4, 2001, pp. 538–543.
Original Russian Text Copyright © 2001 by Golovanov, Tsygankova.

Structure–Property Correlation Equation: IV.[1] Estimation of the Logarithm of the Distribution Factor of Flexible Hydrocarbon Molecules in the Octanol–Water System

I. B. Golovanov and I. G. Tsygankova

Institute of Theoretical and Experimental Biophysics, Russian Academy of Sciences, Pushchino, Moscow oblast, Russia

Received August 16, 1999

Abstract—A simple structure–property correlation equation was suggested for estimating the distribution factors of compounds in the octanol–water system. The ways of taking into account the conformational features of molecules are discussed. The parameters of the correlation equation are determined from a limited set of test examples.

We have described in [1] a method allowing fairly accurate estimation of the distribution factor ($\log P$) in the octanol–water system of various classes of organic compounds with a "stretched" conformation in which interactions between "remote" fragments are insignificant. However, in large molecules, fragments separated by a large distance in the chain can closely approach each other in the space, so that their interaction cannot be neglected; in such cases the estimations according to [1] do not give adequate results. This is due to the lack of the required experimental data and to difficulties in adequate estimation of these interactions, since the integrals describing these interactions include various correlation functions whose estimation in relation to the molecular conformation is impossible without sophisticated calculations.

However, the problem is urgent, and it is necessary to look for ways to solve it. At present, the $\log P$ values for large molecules are approximately estimated assuming that $\log P$ is proportional to the accessible surface area of the molecule [2–6]:

$$\log P = k_1 + kS,$$

where k_1 and k are parameters (constants), and S is the surface area of the molecule accessible for solvent molecules.

A change in the molecular conformation causes a change in the accessible surface area and hence in $\log P$. In going from conformation i to conformation j, the change in $\log P$, $\Delta\log P_{ij}$, will be given by

$$\Delta\log P_{ij} = k(S_i - S_j) = k\Delta S, \tag{1}$$

where k is a parameter, and ΔS is the difference between the accessible surface areas of conformers i and j.

In terms of this approach, the same parameter k should describe both the $\log P$ values for molecules and the $\Delta\log P$ values for various conformational transitions. It would be easier if these quantities were described by independent parameters.

Another possibility is provided by the approach suggested in [1]. Using this approach, we can estimate with a good accuracy $\log P$ for molecules in the *trans* conformation, and then, reasoning in terms of the perturbation theory, we can assume that transition into another conformation is accompanied by a change in the accessible surface area of the molecule and hence in $\log P$ according to Eq. (1). In this case, $\log P$ of the molecule in conformation i will be given by

$$\log P_i = \log P_{trans} + k(S_i - S_{trans}) = \log P_{trans} + k\Delta S_{i,trans}. \tag{2}$$

Here, $\Delta\log P_{ij}$ will also be determined by Eq. (1), but the value and physicochemical sense of parameter k will be different. In many cases $\Delta\log P_{ij}$ is much less than $\log P_i$ and $\log P_j$, and even if it will be determined with a large error, the values of $\log P_i$ and $\log P_j$ will be considerably more accurate.

---

[1] For communication III, see [1].

Table 1. Values of $\log P_{1...4}$, $\Delta S_{1...4}$, and parameter k for model benzene derivatives

Molecule	$\log P$ [7]	Expression for $\log P$ [Eq. (3)]	$m\Delta\log P_{1...4}$	S, Å^2	$\Delta S_{1...4}$, Å^2	k
1,4-Dimethylbenzene	3.15	A		310.2		
1,2-Dimethylbenzene	3.12	$A + \log P_{1...4}$	–0.03	302.1	–8.1	0.0037
1,3,5-Trimethylbenzene	3.42	B		339.9		
1,2,3-Trimethylbenzene	3.61	$B + 2\ \log P_{1...4}$	+0.19 [a]	324.5	–15.5	
1,2,4,5-Tetramethylbenzene	4.00	C		355.2		
1,2,3,4-Tetramethylbenzene	3.98	$C + \log P_{1...4}$	–0.02	347.4	–7.3	0.0027
Pentamethylbenzene [b]	(4.58)	D		401.4		
Pentamethylbenzene	4.56	$D + 4\ \log P_{1...4}$	–0.02	370.4	–31	0.0007
Hexamethylbenzene [b]	(5.07)	E		431.9		
Hexamethylbenzene	4.61	$E + 6\ \log P_{1...4}$	–0.46	385.5	–46.4	0.010
1,4-Dimethylnaphthalene	4.37	F		359.3		
1,8-Dimethylnaphthalene	4.26	$F + \log P_{1...5}$	–0.11	349.4	–9.9	0.011

[a] This may be an erroneous result. It is obvious, from the general knowledge, that $\log P$ of 1,2,3-trimethylbenzene should be lower than that of 1,3,5-trimethylbenzene. [b] $\log P$ was calculated by Eq. (3) assuming no interactions between CH_3 groups ($m = 0$).

Here we demonstrate the ways of choosing k and the potential of our approach.

Apparently, the solution of the problem involves consideration of a series of previously studied molecules having different conformations, for each of which the $\log P$ values would be known from the experiment or their accurate estimates would be available. Unfortunately, such data are very few. Actually, we consider below all the examples known by us. In what follows, we used the experimental data on $\log P$ from [7]. The accessible surface areas were calculated by programs based on the method in [8].

The simplest examples of molecular conformers differing in the relative positions of the CH_3 groups are benzene and naphthalene derivatives. Apparently, $\log P$ for an n-methyl benzene derivative can be approximately estimated as

$$\log P_{(CH_3)_nC_6H_{6-n}} \approx \log P_{C_6H_6} + n\log P_{CH_3} + m\log P_{1...4}, \quad (3)$$

where $\log P_{C_6H_6}$ is the logarithm of the distribution factor of benzene in the octanol–water system; $\log P_{CH_3}$ is the contribution from introduction of one CH_3 group; n is the number of CH_3 groups; $\log P_{1...4}$ is the contribution from interaction of two CH_3 groups in the o-position (1...4 interaction). For CH_3 groups in the m- and p-positions, such contributions are taken as zero (no interaction between CH_3 groups).

The $\log P_{CH_3}$ values are determined from the experimental data for benzene, toluene (one noninteracting CH_3 group), 1,4-dimethylbenzene (two noninteracting Ch_3 groups), and 1,3,5-trimethylbenzene (three noninteracting CH_3 groups). The average value of $\log P_{CH_3}$ estimated from these data is $+0.49$.

In some cases the contribution $\log P_{1...4}$ can be determined from experimental data. For example, the difference between $\log P$ of 1,2-dimethylbenzene and 1,4-dimethylbenzene gives $\log P_{1...4}$, etc. However, to estimate $\log P_{1...4}$ from data for hexamethylbenzene, it is necessary to calculate $\log P$ for the case if 1...4 interactions were absent. This can be done using Eq. (3), assuming $m = 0$.

The resulting values of $\log P_{1...4}$ corresponding to interaction of o-methyl groups and the accessible surface areas and parameters k of benzene derivatives are listed in Table 1.

It is seen that the calculated values of $\log P_{1...4}$ are the differences between the close values of $\log P_i$; $\log P_i$ is determined with an error of ±(0.01–0.10), and the errors in determination of $\log P_{1...4}$ exceed this level. Therefore, determination of k from these data is impossible. It is necessary to find examples in which the effect from the mutual approach of groups would exceed 0.2. Such effects are observed in methyl ethyl and diethyl benzene derivatives (Table 2).

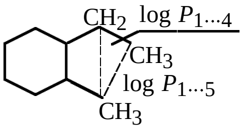
In these cases, the effect from mutual approach of the groups exceeds the error in experimental determination of $\log P$ of different conformers, so that the resulting values of k are reasonable. Apparently, the second value of k obtained from data for diethylbenzene is more probable.

The effects from mutual approach of the CH_3 and CH_2 groups can be estimated from comparison of the

Table 2. Values of log *P*, accessible surface areas, and coefficient *k* for methyl ethyl and diethyl benzene derivatives

Molecule	log *P* [7]	Expression for log *P* [Eq. (3)]	$\Sigma \log P_{ij} = \Delta \log P$	*S*, Å^2	Δ*S*, Å^2	*k*
1,4-Methylethylbenzene	3.90	A		336.7		
1,2-Methylethylbenzene	3.53	$A + \log P_{1...4} + \log P_{1...5}{}^{a}$	–0.37	326.7	–10	0.037
1,4-Diethylbenzene	4.45	B		362.4		
1,2-Diethylbenzene	3.72	$B + \log P_{1...4} + 2 \log P_{1...5}{}^{a} + {}$ $+ \log P_{1...6}{}^{a}$	–0.73	347.6	–15.3	0.048

[a] Contributions from interaction of remote groups:



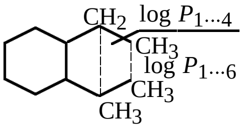
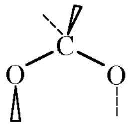


Table 3. Values of log *P*, accessible surface areas, and distribution factors of molecules isoelectronic with C_5 and C_7

Molecule	log *P* [7]	Δlog *P*	ΔΔlog *P*	*S*, Å^2	Δ*S*, Å^2	ΔΔ*S*, Å^2	*k* [a]
n-C_5H_{12}	3.39			278			
n-C_7H_{16}	4.66	1.27		338	60		
$CH_3OCH_2OCH_3$	0.18			257			
$C_2H_5OCH_2OC_2H_5$	0.84	0.66	–0.61	309	52	–8	0.082

[a] $k = \Delta\Delta \log P / \Delta\Delta S$.

log *P* values for molecules whose preferable structures are close to the trans conformations and for those in which the middle sections have the gauche–gauche conformation resulting in mutual approach of the terminal groups. The chain lengthening in such molecules should lead to different effects. For example, in going from pentane to heptane and from diethyl to dipropyl ether, log *P* changes by 1.27 and 1.14, respectively. However, in going from dimethoxymethane $CH_3OCH_2OCH_3$ to diethoxymethane $C_2H_5OCH_2 \cdot OC_2H_5$ (molecules isoelectronic with pentane and heptane), log *P* changes by only 0.66. This difference between acetals, on the one hand, and hydrocarbons and ethers, on the other hand, is due to the fact that the anomeric effect produced by repulsion of the long electron pairs of the O atoms causes the acetal to take the following conformation:



In this conformation, the terminal groups, in contrast to pentane and heptane, approach each other in space.

Data on such molecules, from which it is possible to estimate *k*, are listed in Table 3.

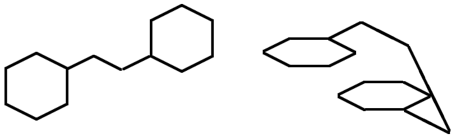
The resulting value of *k* appeared to be larger than that determined in the previous examples, which is probably due to the fact that in calculations we considered only the major conformers, whereas the other conformers, whose content in the equilibrium mixture could be significant, were neglected.

Apparently, such effects can be found in molecules of the type $R(CH_2)_nR$, where R is a bulky hydrophobic group, e.g., $CH{=}CH_2$, $C(CH_3)_3$, C_6H_5, and *n* takes the values from 1 to 5. However, from this series, we found data only for three benzene derivatives (Table 4).

In going from diphenylmethane to dibenzyl and 1,3-diphenylpropane, the number of CH_2 groups in the molecule successively increases, and, assuming no interaction between the phenyl rings, we should expect increase in log *P* along this series. However, 1,3-diphenylpropane actually exhibits the lowest log *P*.

Whereas in the preferable conformation of dibenzyl the phenyl rings are remote from each other and no interaction between them can be assumed, in the

1,3-diphenylpropane molecule the phenyl rings approach each other:



The accessible surface area of this conformation, as compared to the stretched conformation, decreases by 23.8 Å. Specifically mutual approach of the phenyl rings gives rise to the contribution $P_{Ph\cdots Ph}$ from interaction of the phenyl rings. From the experimental data and calculation assuming no interaction, we obtain $P_{Ph\cdots Ph} \cong 3.43 - 5.18 = -1.75$, and hence $k = 1.75/23.8 \approx 0.072$.

From the above examples, we obtained four, significantly different values of k (0.037, 0.048, 0.082, and 0.072). However, at present it is impossible to assess their reliability; therefore, we will suggest the mean value of 0.06. It should be noted that in all the above reasonings it was implicitly assumed that the molecules in question occur either in trans conformation or in a certain more compact conformation. Actually, however, not only the most preferable but also some other similar conformations may be more compact, which will result in overestimated ΔS and hence underestimated k. Along with the above-discussed data, only experimental data for linear C_{12} and C_{14} hydrocarbons are available to us [7]. Their $\log P$ values calculated according to [1] exceed the experimental values by 0.95 and 0.89, respectively.

The decrease in $\log P$ is due to the presence in the equilibrium mixture of conformers with the *gauche* configuration of some fragments. Indeed, each of these molecule has numerous conformers only slightly differing in energy but strongly differing in the relative content; without estimation of their relative content, it is impossible to estimate ΔS. For this purpose, we performed the conformational and molecular-dynamics calculations of C_{12} and C_{14} molecules using the programs developed by N.K. Balabaev (Institute of Mathematical Problems of Biology) [9]. Data on the relative energies, relative contents, and accessible surface areas of different conformations of $C_{12}H_{26}$ are listed in Table 5.

Table 4. Values of $\log P$ for some benzene derivatives

Molecule	$\log P$ [7] (exp.)	$\log P$ [1] (calc.)	Expression for $\log P$ [1]
$C_6H_5CH_2C_6H_5$	4.14	4.06	A
$C_6H_5(CH_2)_2C_6H_5$	4.79	4.59	$A + \alpha + \beta + 2\gamma$
$C_6H_5(CH_2)_3C_6H_5$	3.43	$5.18 + \log P_{Ph...Ph}$	$A + 2\alpha + 2\beta + 2\gamma + \log P_{Ph...Ph}$

Table 5. Characteristics of some conformers of $C_{12}H_{26}$

Conformer [a]	Relative energy, kcal mol^{-1}	S, Å^2	Relative content, $\times 10^{-2}$
ttt...t	0	488.5	0.17
1 *gauche*	~0.6	~483.5	1.6
2 *gauche*	~1.0	~479.5	4.6
3 *gauche*	~1.5	~472.3	2.5
$(tgg)_3$	5.26	411.3	>0.01
“Pin”	2.53	415.1	

[a] 1 *gauche* means that one of torsion angles corresponds to the *gauche* conformation of the fragment; 2 *gauche* means that there are two torsion angles corresponding to such a conformation; “pin” is a ⊂-like conformation.

The most populated are the states in which there is one chain section in the *gauche* state with the other sections being in the *trans* state (1 *gauche*), two chain sections in the *gauche* state separated by sections in the *trans* state (2 *gauche*), etc. The other states are significantly less populated. Therefore, for approximate estimations of the variation of the accessible surface areas of C_{12} and C_{14} we assumed that these compounds exist as an equilibrium mixture of 1 *gauche*, 2 *gauche*, and 3 *gauche* conformers with equal relative contents. From data in Table 5, we obtain

$$\Delta S \approx -(5 + 9 + 16.2)/3 \text{ Å}^2 = -10 \text{ Å}^2.$$

Such a decrease in ΔS corresponds to a decrease in $\log P$ by at least 0.6. With C_{14}, the pattern (and hence $\Delta\log P$) will be similar. As a result, estimation of $\log P$ for C_{12} and C_{14} gives 6.45 and 7.49, which is close to the experimental data (6.10 and 7.20, respectively). This agreement is quite reasonable taking into account the rough approximations made.

These examples show that in terms of our approach [based on relation (2)] it is possible to obtain results qualitatively consistent with the experiment, even when ΔS is determined as the difference between the accessible surface areas of the trans conformer and the preferable conformer. The latter can be chosen on the basis of general knowledge, without conformational analysis. However, for more accurate estimations it is necessary to use the methods of molecular dynamics,

Table 6. Accessible surface areas of hydrocarbon molecules

Molecule	Relative energy of conformer, kcal mol^{-1}	S, Å^2
1,4-Diethylbenzene	0 [a]	485.4
1,2-Diethylbenzene	0 [a]	437.3
	0.04	457.8
	0.42	470.1
	0.72	451.6
1,4-Dipropylanthracene	0 [a]	596.9
1,8-Dipropylanthracene	0 [a]	523.0
	0.06	533.2
	0.48	520.8
1,4-Dibutylanthracene	0 [a]	657.0
1,8-Dibutylanthracene	0 [a]	544.4
	0.25	574.8
1,4-Diphenylbutane	0 (*trans*)	482.6
	–0.72 [a]	461.0
	0.47	427.0
1,5-Diphenylpentane	0 (*trans*)	511.8
	–8.46 [a]	459.8

[a] Preferable conformation.

since they allow prediction of the set of the conformers corresponding to the equilibrium mixture.

The suggested value of k was chosen from a limited set of available experimental data. The more substantiated conclusion on the potential of the method and more accurate estimations of k can be made after performing a series of experiments with compounds in which mutual approach of remote groups can cause significant changes in $\log P$ and S. We considered preliminarily a number of systems, including conformational analysis and calculation of accessible surface areas of conformers corresponding to low-energy local minima in which the desired effects could be manifested. Some of the examples are given in Table 6. It is seen that in the first example ΔS in going from 1,4-diethylbenzene to 1,2-diethylbenzene is about 30 Å^2 (for 1,2-diethylbenzene, S was taken as average of the conformers given in Table 6); in going from 1,4-dipropylanthracene to 1,8-dipropylanthracene, ΔS is about 70 Å^2, etc., which causes significant differences in $\log P$ as compared to the values calculated assuming no interaction between the radicals. Diphenyl derivatives of alkanes are one more example. Assuming that the conformations given in Table 6 are present in the equilibrium mixture with equal weights, we obtain that the accessible surface area of 1,4-diphenylbutane is 457 Å^2, whereas the accessible surface area of the *trans* conformer in which the phenyl rings do not interact is 482 Å^2. ΔS is about 25 Å^2; therefore, the difference between the experimental value of $\log P$ and the value calculated assuming no interaction of phenyl rings should be significant. In 1,5-diphenylpentane, the preferable conformer significantly differs in energy from the *trans* conformer owing to favorable interaction of phenyl rings. Their approach decreases the accessible surface area by ~50 Å^2; as a result, $\log P$ considerably decreases. Also, we can note cyclic peptides consisting of an even number of amino acid residues. If a cyclic peptide consists of alternating *d* and *l* amino acids, their pendant groups will be oriented outside of the ring, and if it consists of *l* amino acids, a half of the pendant groups will be oriented inside the ring, which will cause a decrease in S and $\log P$. Measurement of the distribution factors of such molecules will allow fairly accurate estimations of k and assessment of the potential of our approach.

REFERENCES

1. Golovanov, I.B. and Tsygankova, I.G., *Zh. Obshch. Khim.*, 2000, vol. 70, no. 2, pp. 223–230.
2. Hermann, R.B., *J. Phys. Chem.*, 1972, vol. 76, no. 19, pp. 2754–2775.
3. Hermann, R.B., *J. Phys. Chem.*, 1975, vol. 79, no. 2, pp. 163–169.
4. Chotia, C., *Nature*, 1974, vol. 248, no. 5446, pp. 338–339.
5. Reynolds, J.A., Gilbert, D.B., and Tanford, C., *Proc. Nat. Acad. Sci. USA*, 1974, vol. 71, no. 8, pp. 2925–2927.
6. Eisenberg, D. and McLachlan, A.D., *Nature*, 1986, vol. 319, no. 6050, pp. 199–203.
7. Hansch, C., Leo, A., and Hoekman, D., *Exploring QSAR*, Washington: Am. Chem. Soc., 1995, vol. 2.
8. Lee, B. and Richards, F.M., *J. Mol. Biol.*, 1971, vol. 55, no. 3, pp. 379–400.
9. Balabaev, N.K., in *Metod molekulyarnoi dinamiki v fizicheskoi khimii* (Method of Molecular Dynamics in Physical Chemistry), Moscow: Nauka, 1996, pp. 258–279.