QSPR MODELING OF THE OCTANOL/WATER PARTITION COEFFICIENT OF ALCOHOLS BY MEANS OF OPTIMIZATION OF CORRELATION WEIGHTS OF LOCAL GRAPH INVARIANTS

Duchowicz[1], P.R.; Castro[1], E. A.; Toropov[2], A. A.; Nesterova[2], A. I.; Nabiev[2], O. M.

[1] INIFTA, Departamento de Química, Facultad de Ciencias Exactas, UNLP,
Suc. 4, C.C. 16, La Plata 1900, Argentina
FAX: +54 221 425 4642; direccion@inifta.unlp.edu.ar
[2] Uzbekistan Academy of Sciences, Algorithm-Engineering Institute, F. Khodjaev Street 25, 100125 Tashkent, Uzbekistan

Received November 10th, 2003. In final form February 23rd, 2004
Dedicated to Prof. Pedro J. Aymonino on the occasion of his 75th birthday

Abstract

A particular approach based on the concept of flexible topological descriptors,the so called "Optimization of Correlation Weights of Local Graph Invariants",is applied to model the octanol/water partition coefficient of a representative setof 62 alcohols. Predictions are quite satisfactory and the numerical data improve previous results based on the application of a novel atomic-level-based AI topological descriptor. Some possible further extensions of the method are pointed out.

Resumen

Se aplica una aproximación particular basada en el concepto de descriptores topológicos flexibles, la denominada "Optimización de los Pesos de Correlación de Invariantes de Grafos Locales" al modelado del coeficiente de partición octanol/agua para un conjunto representativos de 2 alcoholes. Las predicciones son bastante satisfactorias y los resultados numéricos constituyen una mejora en resultados previos basados en la aplicación de un nuevo descriptor topológico, el denominado descriptor de nivel atómico de base AI. Finalmente, se destacan algunas posibles extensiones del método.

Introduction

The progress in computer technology during the last 25 years has enabled the performance of ever more precise quantum mechanical calculations related to structure and interactions of chemical compounds. However, the qualitative models relating electronic structure to molecular geometry have not progressed at the same pace. There is a continuing need in chemistry for simple concepts and qualitatively clear pictures that be also quantitatively comparable to *ab initio* quantum chemical calculations. Topological methods, and, more specifically, graph theory as s fixed-point topology, provide in principle a chance to fill this gap [1]. With more than 100 years of application to chemistry, graph theory has proven to be of vital importance as the most natural language of chemistry. The explosive development of chemical graph theory during the

last 30 years has increasingly overlapped with quantum chemistry. Besides contributing to the solution of various problems in theoretical chemistry, this development indicates that topology is an underlying principle that explains the success of quantum mechanics an goes beyond it, thus promising to bear more fruit in the future.

Most applications of data analysis involve attempts to fit a model, usually quantitative, to a set of experimental measurements or observations. The reasons for fitting such models are varied. For example, the model may be purely empirical and be required in order to make predictions for new experiments. On the other hand, the model may be based on some theory or law, and an evaluation of the fit of the data to the model may be used to give insight into the process underlying the observations made. In some cases the ability to fit a model to a set of data successfully may provide the inspiration to formulate some new hypothesis. The type of model which may be fitted to any set of data depends not only on the nature of the data but also on the intended use of the model. In many applications a model is meant to be used predictively, but the predictions need not necessarily be quantitative [2].

The majority of molecular discoveries today are the result of an iterative, three-phase cycle of design, synthesis and test. Analysis of the results from one iteration provides information and knowledge that enables the next cycle to be initiated and further improvements to be achieved. A common feature of this analysis stage is the construction of some form of model which enables the observed activity or properties to be related to the molecular structure. Many types of models are possible, with mathematical and statistical models being particularly common. Such models are often referred to as Quantitative Structure-Activity Relationships (QSAR) or Quantitative Structure-Property Relationships (QSPR).

The basic principle of QSAR/QSPR theory is the mathematical relationship

$$p = f(s) \qquad (1)$$

where p is any biological activity or physicochemical property, s is a set of variables associated to the molecular structure (they are called *molecular descriptors*) and f is an arbitrary function. Molecular descriptors are numerical values that characterize properties of molecules. For example, they may represent the physicochemical properties of a molecule or they may be values that are derived by applying algorithmic techniques to the molecular structure. Many different molecular descriptors have been described and used for a wide variety of purposes. They vary in the complexity of the information they encode and in the time required to calculate them.

The biochemical interactions in the living cell occur in both aqueous and hydrophobic media (*i.e.* coupling to an active site of an enzyme, transport through a biomembrane) [3]. In addition, other pharmaco-kinetic properties are related to the difference in solubility of bioactive molecules in aqueous and organic solvents. Hence, it is important to account properly for the solute interactions in both aqueous and organic media. The partition of chemical compounds between organic and aqueous phases is often modeled by the octanol/water partition coefficient (log P) [4], because it is assumed that octanol may reflect lipid tissues in living organisms. Log P has been successfully related to bioconcentration factor, soil and sediment sorbtion partition coefficients and to toxicities of organic chemicals towards aquatic organisms. Direct measurement of P by means of the shake-flask procedure yields only reliable data for chemicals with log P less

than 4-5 [5]. P of more hydrophobic substances can be measured either by the generator-column-method or by the slow-stirring technique. In addition to these direct approaches to the determination of P, several other methods were employed: (1) calculation based on molecular fragments additivity [6,7], (2) correlation with capacity factors on reversed-phase HPLC, (3) correlation with molecular descriptors (volume, surface area, molar refraction, parachor, molecular weight) [8], and (4) correlation with molar volume and solvachromic parameters.

The partitioning of a hydrophobic solute between octanol and water is due to the difference between the interactions that the solute is experiencing in water versus octanol. Hence, the relationship between water solubility and P [6] has been studied extensively over the last two decades [9]. Examples of such correlation have been published for halogenated benzene, aromatic hydrocarbons, aldehydes, esters and alcohols. However, due to experimental difficulties, few accurate data for compounds with log P greater than 6 have been reported, which limits the use of such correlation for prediction purposes.

In a series of rather recent studies, Ren [10-12] derived a new atom-type AI topological indices from the adjacency matrix and distance matrix of a graph to model six properties of alkanes. Further, high quality models were developed to correlate four physical properties of a small data of alcohols and three physical properties of a mixed set of compounds containing alkanes and alcohols with their structures. The atom-type AI indices offer the possibility of understanding the role of individual groups in molecules. In a latter paper, Ren [13] have illustrated the application of the novel AI indices to a wide range of physical properties and especially biological activities that depend on the strength of intermolecular interactions such as hydrogen bonding interactions of –OH moieties in molecules. The author calculated the octanol/water partition for 62 alcohols via a multiple linear regression to develop the structure-property model based on the modified Xu (X_u^m) and AI indices. The best two-parameter model show that although X_u^m makes a major contribution to octanol/water partition, which indicates the additive behavior of the property, other atomic groups, especially –OH groups, are also important factors influencing the values of this property.

Since there are other alternatives to predict log P within the frame of the QSPR theory, we have deemed sensible to look for ways to improve these predictions. An interesting and very promising option is the approach based upon the correlation weights of local graph invariants [14-16], which has proved to be a quite suitable tool to calculate thermodynamic properties for a wide variety of molecular species [17-22].

The paper is organized as follows: the next section deals with the presentation of the method and the mathematical algorithm applied in this study. Then, we display the set of alcohols together with available experimental data and previous theoretical prediction of partition coefficient plus the results derived from the present approach and they are discussed in a comparative fashion. Finally, we analyze the main conclusions derived from this study and point out some possible further extensions of this calculation method.

Method

Molecular descriptors employed in QSPR theory can be divided into two broad categories: fixes variable descriptors. Fixed descriptors are molecular invariants that can be numerically computed once a molecule is selected. This is the case with the great majority of proposed

hundreds pf descriptors. Variable descriptors involve one or more variables, the values of which are selected during the regression process. Hence, a variable descriptor can either be a function of a single variable or function of several variables.

In contrast to the traditional molecular indices, which one can calculate after selecting a set of compounds to be studied and then proceed with statistical analysis, the variable indices are initially non-numerical. Therefore, they cannot be calculated in advance for the set of compounds. Instead, one starts with an arbitrary set of values for the yet undetermined variables and, in an iterative procedure, varies these initial values seeking values that will produce the smallest standard error of the property under consideration. It is clear that the use of variable (they are also called flexible) descriptors can only improve correlations over the use of simple indices because, if all the variables took on a zero value (which is very unlikely), we would obtain the results that coincide with the results based on the traditional molecular indices.

Among the several existing options to employ flexible molecular descriptors, the optimization of Correlation Weights of Local Graph Invariants (OCWLI) has shown to be a suitable possibility to employ in QSPR theory and results have been very encouraging [14-22]. The method has been described in detail in the current literature so that we do not deem necessary to repeat it here. The interested reader can consult the pertinent bibliography [14-22].

Regarding the choice of the f function in relationship (1), we have pointed out that it is arbitrary. The simplest mathematical structure is the linear one, i.e.

$$p = A + Bs, \qquad (2)$$

where A and B are two numerical coefficients to be determined by a standard least square criterion and s stands for a single molecular descriptor. In this work we have resorted to this linear relationship since it provides good enough results, so that the employment of other more complex formulae would not improved significantly the final fittings. Mathematical software employed here is the well known MATHEMATICA® computer program [23].

The total molecular set of 62 alcohols is the same as that employed by Ren [13]. We have employed two numerical approaches: a) The calculations were made on the complete set, and b) The calculations were made in two subsets. In this second case, we have divided the complete set into two subsets for calculations: a training set and a test set comprising 31 alcohols each one. In order to test whether the choice of molecules in each set influences the final results, we have performed several choices, and we have seen they furnish practically the same results, so that we report data for a representative partition. Obviously, results for the second set are true predictions.

Results And Discussion

In Table 1 we present the main statistical characteristics of the OCWLI models.

Table 1. Statistical characteristics of the OCWLI models

Test set

n	R	S	F
31	0.9893	0.222	1329
31	0.9893	0.222	1329
31	0.9893	0.222	1329
31	0.9970	0.133	4747
31	0.9970	0.133	4747
31	0.9969	0.134	4686
31	0.9953	0.190	3077
31	0.9957	0.178	3388
31	0.9955	0.187	3196

Complete set

n	R	s	F
62	0.9914	0.214	3432
62	0.9914	0.214	3432
62	0.9914	0.214	3432
62	0.9973	0.124	10989
62	0.9973	0.124	10985
62	0.9973	0.125	10908
62	0.9966	0.148	8798
62	0.9969	0.141	9551
62	0.9967	0.146	9014

LIs denotes local LFFG invariants

NLIs denotes the number of parameters of the OCW

EC0, EC1, EC2 are Morgan extended connectivity indices [24] of zero, first and second order, respectively.

The models under consideration to calculate octanol/water partition coefficients are:

$$\log P = A\ DCW(a,ECX) + B \qquad (3)$$

$$DCW(a,ECX) = \Sigma [\ CW(a)\ CW(ECX)\ ] \qquad (4)$$

In Tables 2-4 we display the CW data for the three probes.

Table 2. Numerical values of the CWs on DCW(a,EC0)

LHFG invariant	CWs of probe 1	CWs of probe 2	CWs of probe 3
C	2.393	2.599	2.393
H	2.351	2.488	2.295
O	3.261	2.351	2.433
0001	2.487	2.613	2.874
0002	2.730	2.591	3.011
0004	2.426	2.334	2.426

Table 3. Numerical values of the CWs on DCW(a,EC1)

LHFG invariant	CWs of probe 1	CWs of probe 2	CWs of probe 3
C	2.091	2.048	2.281
H	1.114	0.587	0.500
O	6.192	5.108	6.192
0002	6.192	6.811	6.192
0004	0.525	0.997	0.637
0005	6.192	5.108	7.430
0007	0.912	0.775	0.812
0008	1.851	1.771	1.418
0010	1.238	1.069	1.154
0011	1.697	1.624	1.379
0013	1.353	1.175	1.336
0014	0.550	0.575	0.537
0016	1.287	1.113	1.364

The final fitting equations based on OCWLI for the octanol/water partition coefficient are the following

$$\log P = 0.03137\ DCW(a,ec0) - 2.268 \quad (5)$$

$$\log P = 0.1422\ DCW(a,ec1) - 7.954 \quad (6)$$

$$\log P = 0.2000\ DCW(a,ec2) - 3.127 \quad (7)$$

Table 4. Numerical values of the CWs on DCW(a,EC2)

LHFG invariant	CWs of probe 1	CWs of probe 2	CWs of probe 3
C	0.987	1.076	1.532
H	0.697	0.475	0.525
O	2.151	1.861	2.228
0005	8.916	4.300	18.488
0007	0.912	0.887	0.825
0008	1.617	1.628	1.474
0010	1.125	1.170	1.113
0011	1.603	1.360	1.362
0013	1.172	1.317	1.222
0016	0.475	0.575	0.512
0020	0.637	0.812	0.662
0022	1.263	1.303	1.221
0023	0.825	0.898	0.762
0025	0.991	1.069	1.016

Table 4 continuing

0026	0.775	0.838	0.725
0028	1.078	1.098	1.073
0029	0.898	0.875	0.787
0031	0.941	1.148	0.991
0032	0.362	0.525	0.425
0034	1.805	1.717	1.591
0035	1.869	1.611	1.451

Calculations with Eqs. (5) – (7) are displayed in Tables 5 – 10.

Table 5. Log P Model based on DCW(a,EC0) – training set.

n	Molecule	DCW	exp.	calc.	exp.-calc.
1	ethanol	55.595	-0.31	-0.52	0.21
2	2-propanol	73.094	0.05	0.03	0.03
3	2-methyl-1-propanol	90.594	0.65	0.57	0.08
4	1-pentanol	108.093	1.40	1.12	0.28
5	2-pentanol	108.093	1.14	1.12	0.02
6	3-pentanol	108.093	1.14	1.12	0.02
7	2-methyl-2-butanol	108.093	0.89	1.12	-0.23
8	1-hexanol	125.592	2.03	1.67	0.36
9	3-hexanol	125.592	1.61	1.67	-0.06
10	2-ethyl-1-butanol	125.592	1.78	1.67	0.11
11	4-methyl-2-pentanol	125.592	1.67	1.67	0.00
12	3,3-dimethyl-1-butanol	125.592	1.57	1.67	-0.10
13	2,2-dimethyl-1-butanol	125.592	1.57	1.67	-0.10
14	3,3-dimethyl-2-butanol	125.592	1.19	1.67	-0.48
15	1-heptanol	143.091	2.34	2.22	0.12
16	3-heptanol	143.091	2.31	2.22	0.09
17	2,2-dimethyl-1-pentanol	143.091	2.39	2.22	0.17
18	4,4-dimethyl-1-pentanol	143.091	2.39	2.22	0.17
19	2,4-dimethyl-1-pentanol	143.091	2.19	2.22	-0.03
20	2,4-methyl-2-pentanol	143.091	1.67	2.22	-0.55
21	2,4-dimethyl-3-pentanol	143.091	2.31	2.22	0.09
22	2,3-dimethyl-3-pentanol	143.091	1.67	2.22	0.55
23	2,2-dimethyl-3-pentanol	143.091	2.27	2.22	0.05
24	3-nonanol	178.090	3.36	3.32	0.04
25	4-nonanol	178.090	3.36	3.32	0.04
26	5-nonanol	178.090	3.36	3.32	0.04
27	1-decanol	195.589	4.01	3.87	0.14
28	1-undecanol	213.089	4.42	4.42	0.00
29	1-tetradecanol	265.586	6.11	6.06	0.05
30	1-pentadecanol	283.086	6.64	6.61	0.03
31	1-hexadecanol	300.585	7.17	7.16	0.01

Average absolute deviation = 0.14

Table 6. Log P Model based on DCW(a,EC0) – test set.

n	Molecule	DCW	exp.	calc.	exp.-calc.
1	1-propanol	73.094	0.34	0.03	0.32
2	1-butanol	90.594	0.84	0.57	0.27
3	2-butanol	90.594	0.61	0.57	0.04
4	2-methyl-2-propanol	90.594	0.37	0.57	-0.20
5	3-methyl-1-butanol	108.093	1.42	1.12	0.30
6	2-methyl-1-butanol	108.093	1.14	1.12	0.02
7	3-methyl-2-butanol	108.093	1.14	1.12	0.02
8	2,2-dimethyl-1-propanol	108.093	1.36	1.12	0.24
9	4-methyl-1-pentanol	125.592	1.78	1.67	0.11
10	2-hexanol	125.592	1.61	1.67	-0.06
11	2-methyl-1-pentanol	125.592	1.78	1.67	0.11
12	2-methyl-2-pentanol	125.592	1.39	1.67	-0.28
13	3-methyl-2-pentanol	125.592	1.67	1.67	0.00
14	2-methyl-3-pentanol	125.592	1.67	1.67	0.00
15	3-methyl-3-pentanol	125.592	1.39	1.67	-0.28
16	2,3-dimethyl-2-butanol	125.592	1.17	1.67	-0.50
17	4-heptanol	143.091	2.31	2.22	0.09
18	5-methyl-2-hexanol	143.091	2.19	2.22	-0.03
19	2-methyl-3-hexanol	143.091	2.19	2.22	-0.03
20	2-methyl-2-hexanol	143.091	1.84	2.22	-0.38
21	3-methyl-3-hexanol	143.091	1.87	2.22	-0.35
22	3-ethyl-3-pentanol	143.091	1.87	2.22	-0.35
23	2,3-dimethyl-2-pentanol	143.091	2.27	2.22	0.05
24	1-octanol	160.591	3.15	2.77	0.38
25	2-octanol	160.591	2.84	2.77	0.07
26	2-ethyl-1-hexanol	160.591	2.84	2.77	0.07
27	1-nonanol	178.090	3.57	3.32	0.25
28	2-nonanol	178.090	3.36	3.32	0.04
29	2,6-dimethyl-4-heptanol	178.090	3.13	3.32	-0.19
30	1-dodecanol	230.588	5.13	4.97	0.16
31	1-octadecanol	335.584	8.22	8.26	-0.04

Average absolute deviation = 0.17

Table 7. Log P Model based on DCW(a,EC1) – training set.

n	Molecule	DCW	exp.	calc	exp.-calc.
1	ethanol	53.940	-0.31	-0.28	-0.03
2	2-propanol	56.695	0.05	0.11	-0.06
3	2-methyl-1-propanol	61.016	0.65	0.72	-0.07
4	1-pentanol	65.216	1.40	1.32	0.08
5	2-pentanol	64.212	1.14	1.18	-0.04
6	3-pentanol	64.212	1.14	1.18	-0.04
7	2-methyl-2-butanol	61.132	0.89	0.74	0.15
8	1-hexanol	68.974	2.03	1.85	0.18
9	3-hexanol	67.970	1.61	1.71	-0.10
10	2-ethyl-1-butanol	68.533	1.78	1.79	-0.01
11	4-methyl-2-pentanol	67.529	1.67	1.65	0.02
12	3,3-dimethyl-1-butanol	67.713	1.57	1.68	-0.11
13	2,2-dimethyl-1-butanol	67.713	1.57	1.68	-0.11
14	3,3-dimethyl-2-butanol	66.709	1.19	1.53	-0.34
15	1-heptanol	72.732	2.34	2.39	-0.05
16	3-heptanol	71.729	2.31	2.25	0.06
17	2,2-dimethyl-1-pentanol	71.471	2.39	2.21	0.18
18	4,4-dimethyl-1-pentanol	71.471	2.39	2.21	0.18
19	2,4-dimethyl-1-pentanol	71.850	2.19	2.26	-0.07
20	2,4-methyl-2-pentanol	68.207	1.67	1.75	-0.08
21	2,4-dimethyl-3-pentanol	70.846	2.31	2.12	0.19
22	2,3-dimethyl-3-pentanol	68.207	1.67	1.75	-0.08
23	2,2-dimethyl-3-pentanol	70.468	2.27	2.07	0.20
24	3-nonanol	79.245	3.36	3.32	0.05
25	4-nonanol	79.245	3.36	3.32	0.05
26	5-nonanol	79.245	3.36	3.32	0.05
27	1-decanol	84.007	4.01	3.99	0.02
28	1-undecanol	87.766	4.42	4.53	-0.11
29	1-tetradecanol	99.041	6.11	6.13	-0.02
30	1-pentadecanol	102.799	6.64	6.66	-0.02
31	1-hexadecanol	106.557	7.17	7.20	-0.03

Average absolute deviation = 0.09

Table 8. Log P Model based on DCW(a,EC1) – test set.

n	Molecule	DCW	exp.	calc.	exp.-calc.
1	1-propanol	57.699	0.34	0.25	0.09
2	1-butanol	61.457	0.84	0.79	0.06
3	2-butanol	60.453	0.61	0.64	-0.03
4	2-methyl-2-propanol	57.373	0.37	0.20	0.17
5	3-methyl-1-butanol	64.774	1.42	1.26	0.16
6	2-methyl-1-butanol	64.774	1.14	1.26	-0.12
7	3-methyl-2-butanol	63.771	1.14	1.11	0.03
8	2,2-dimethyl-1-propanol	63.955	1.36	1.14	0.22
9	4-methyl-1-pentanol	68.533	1.78	1.79	-0.01
10	2-hexanol	67.970	1.61	1.71	-0.10
11	2-methyl-1-pentanol	68.533	1.78	1.79	-0.01
12	2-methyl-2-pentanol	64.890	1.39	1.27	0.12
13	3-methyl-2-pentanol	67.529	1.67	1.65	0.02
14	2-methyl-3-pentanol	67.529	1.67	1.65	0.02
15	3-methyl-3-pentanol	64.890	1.39	1.27	0.12
16	2,3-dimethyl-2-butanol	64.449	1.17	1.21	-0.04
17	4-heptanol	71.729	2.31	2.25	0.06
18	5-methyl-2-hexanol	71.287	2.19	2.18	0.01
19	2-methyl-3-hexanol	71.287	2.19	2.18	0.01
20	2-methyl-2-hexanol	68.649	1.84	1.81	0.03
21	3-methyl-3-hexanol	68.649	1.87	1.81	0.06
22	3-ethyl-3-pentanol	68.649	1.87	1.81	0.06
23	2,3-dimethyl-2-pentanol	68.207	2.27	1.75	0.53
24	1-octanol	76.491	3.15	2.92	0.23
25	2-octanol	75.487	2.84	2.78	0.06
26	2-ethyl-1-hexanol	76.049	2.84	2.86	-0.02
27	1-nonanol	80.249	3.57	3.46	0.11
28	2-nonanol	79.245	3.36	3.32	0.05
29	2,6-dimethyl-4-heptanol	78.363	3.13	3.19	-0.06
30	1-dodecanol	91.524	5.13	5.06	0.07
31	1-octadecanol	114.074	8.22	8.27	-0.05

Average absolute deviation = 0.09

Table 9. Log P Model based on DCW(a,EC2) – training set.

n	Molecule	DCW	exp.	calc.	exp.-calc.
1	ethanol	14.053	-0.310	-0.32	0.01
2	2-propanol	16.110	0.050	0.10	-0.05
3	2-methyl-1-propanol	19.005	0.650	0.67	-0.02
4	1-pentanol	22.368	1.400	1.35	0.05
5	2-pentanol	21.493	1.140	1.17	-0.03
6	3-pentanol	21.713	1.140	1.22	-0.08
7	2-methyl-2-butanol	19.074	0.890	0.69	0.20
8	1-hexanol	25.000	2.030	1.87	0.16
9	3-hexanol	24.380	1.610	1.75	-0.14
10	2-ethyl-1-butanol	24.399	1.780	1.75	0.03
11	4-methyl-2-pentanol	22.916	1.670	1.46	0.21
12	3,3-dimethyl-1-butanol	23.470	1.570	1.57	0.00
13	2,2-dimethyl-1-butanol	23.725	1.570	1.62	-0.05
14	3,3-dimethyl-2-butanol	22.202	1.190	1.31	-0.12
15	1-heptanol	27.633	2.340	2.40	-0.06
16	3-heptanol	27.013	2.310	2.28	0.03
17	2,2-dimethyl-1-pentanol	27.124	2.390	2.30	0.09
18	4,4-dimethyl-1-pentanol	27.227	2.390	2.32	0.07
19	2,4-dimethyl-1-pentanol	27.243	2.190	2.32	-0.13
20	2,4-methyl-2-pentanol	24.530	1.670	1.78	-0.11
21	2,4-dimethyl-3-pentanol	26.644	2.310	2.20	0.11
22	2,3-dimethyl-3-pentanol	24.305	1.670	1.73	-0.06
23	2,2-dimethyl-3-pentanol	26.454	2.270	2.16	0.11
24	3-nonanol	32.277	3.360	3.33	0.03
25	4-nonanol	32.313	3.360	3.34	0.02
26	5-nonanol	32.313	3.360	3.34	0.02
27	1-decanol	35.529	4.010	3.98	0.03
28	1-undecanol	38.161	4.420	4.51	-0.09
29	1-tetradecanol	46.058	6.110	6.09	0.03
30	1-pentadecanol	48.690	6.640	6.61	0.03
31	1-hexadecanol	51.323	7.170	7.14	0.03

Average absolute deviation = 0.07

Table 10. Log P Model based on DCW(a,EC2) – test set.

n	Molecule	DCW	exp.	calc.	exp.-calc.
1	1-propanol	17.239	0.34	0.32	0.02
2	1-butanol	19.736	0.84	0.82	0.02
3	2-butanol	18.826	0.61	0.64	-0.03
4	2-methyl-2-propanol	16.017	0.37	0.08	0.29
5	3-methyl-1-butanol	21.808	1.42	1.24	0.19
6	2-methyl-1-butanol	22.027	1.14	1.28	-0.14
7	3-methyl-2-butanol	20.934	1.14	1.06	0.08
8	2,2-dimethyl-1-propanol	21.574	1.36	1.19	0.17
9	4-methyl-1-pentanol	24.184	1.78	1.71	0.07
10	2-hexanol	24.126	1.61	1.70	-0.09
11	2-methyl-1-pentanol	24.439	1.78	1.76	0.02
12	2-methyl-2-pentanol	21.092	1.39	1.09	0.30
13	3-methyl-2-pentanol	23.305	1.67	1.53	0.14
14	2-methyl-3-pentanol	23.170	1.67	1.51	0.16
15	3-methyl-3-pentanol	21.482	1.39	1.170	0.22
16	2,3-dimethyl-2-butanol	19.882	1.17	0.85	0.32
17	4-heptanol	27.048	2.31	2.28	0.03
18	5-methyl-2-hexanol	25.942	2.19	2.06	0.13
19	2-methyl-3-hexanol	25.838	2.19	2.04	0.15
20	2-methyl-2-hexanol	23.724	1.84	1.62	0.22
21	3-methyl-3-hexanol	23.499	1.87	1.57	0.30
22	3-ethyl-3-pentanol	25.905	1.87	2.05	-0.18
23	2,3-dimethyl-2-pentanol	24.270	2.27	1.73	0.54
24	1-octanol	30.265	3.15	2.93	0.22
25	2-octanol	29.390	2.84	2.75	0.09
26	2-ethyl-1-hexanol	29.443	2.84	2.76	0.08
27	1-nonanol	32.897	3.57	3.45	0.12
28	2-nonanol	32.022	3.36	3.28	0.08
29	2,6-dimethyl-4-heptanol	29.892	3.13	2.85	0.28
30	1-dodecanol	40.794	5.13	5.03	0.10
31	1-octadecanol	56.587	8.22	8.19	0.03

Average absolute deviation = 0.16

The average absolute deviations for the training and test sets for the different Morgan extended connectivity indices are displayed in Table 11.

Table 11. Average absolute deviations for the different sets.

Descriptor	Training set	Test set	Complete set
DCW(a,EC0) (a)	0.17	0.14	-
DCW(a,EC1) (a)	0.09	0.09	-
DCW(a,EC2) (a)	0.16	0.07	-
AI, X_u^m (b)	-	-	0.12

(a)Present calculation
(b)Ref. 13

The analysis of results presented in Table 11 show several interesting features. The first one is that statistical parameters are nearly the same for the three probes corresponding to each Morgan's index. It means that this approach is consistent (*i.e.* final results are not dependent of the particular probe employed to derive fitting equations). Besides, the overall statistical results are quite satisfactory for the different sets, although those corresponding to the training set are the best ones. Specially important are statistical parameters corresponding to the tests set, since they correspond to real predictive results, while those associated to the training and complete sets are just fitting parameters.

Regarding the behavior of the three Morgan extended indices, the data in Table 1 suggests that EC1 and EC2 are the best ones and this is confirmed when analyzing the absolute average deviations displayed in Table 11. In fact, the average absolute deviations for test sets is 0.09 and 0.07, respectively. These figures deserve to be compared with that corresponding to the Rens' results for the complete set, i.e. 0.12 (see Table 4 in Ref. 13), which show clearly the better quality of present predictions with regard to those published before. In order to judge properly these comparisons, one must take into account that Rens' fitting equation (see Eq. 13 in Ref. 13) depends upon two variables, while present relationships depend on just one variable. Besides, the comparison of statistical coefficients also demonstrate the higher quality of the present equations

Conclusions

We have shown that optimization of correlation weights of local graph invariants are suitable molecular descriptors to model the octanol/water partition coefficients of alcohols. This particular set of flexible topological variables gives quite reliable predictions of this physicochemical property and compares favorably with other recent calculation schemes within the realm of QSPR theory. This finding agrees with other recent similar results for the calculation of other biological activities and physicochemical properties, and it demonstrates the convenience of resorting to variable molecular descriptors for prediction purposes in QSAR/QSPR theory in order to take advantage of this possibility. As stated before, there are other options to improve fitting equations in the regression analysis, such as to employ several variables and/or try different functional algebraic forms for the modeling function f. In this study, it has not been necessary to employ these

resources to get optimal results, but they should not be ignored when applying multilinear regression analysis within the realm of QSAR/QSPR theory.

References

[1] D. Bonchev, O. Mekenyan, Eds., *Graph Theoretical Approaches to Chemical* Reactivity, Kluwer Academic Publishers, Dordrecht, **1994**.

[2] D. Livingstone, *Data Analysis for Chemists*, Oxford Science Publications, Oxford University Press, Oxford, **1995**.

[3] G. L. Zubay, *Biochemistry*, McGraw-Hill, London, **1998**, pp. 3-26.

[4] C. Hansch, A. Leo, *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*, American Chemical Society, Washington, D.C. **1995**, pp.97-168.

[5] E. E. Kenaga, K. A. I. Goring, in *Aquatic Toxicity*, ASTM STP, Vol. 707, J. G. Eaton, P. R. Parrish, A. C. Hendricks (Eds.), American Society for Testing and Materials, **1980**, pp. 78-115.

[6] C. Hansch, A. Leo, *Substituent Constants for Correlation Analysis in Chemistry and Biology*, Wiley-Interscience, New York, **1980**, pp. 9-12.

[7] R. F. Rekker, *The Hydrophobic Fragmental Constants, its Derivation and Application: a Means of Characterizing Membrane Systems*, Elsevier, Amsterdam, **1977**, pp. 73-91.

[8] R. S. Pearlman, in *Partition Coefficient, Determination and Estimation*, W. J. Dunn, J. H. Block, R. S. Pearlman, (Eds.), Pergamon, New York, **1986**, pp. 3-20.

[9] R. S. Pearlman, in *Physical Chemical Properties of Drugs*, S. H. Yalkovsky, A. A. Sinkula, S. C. Valvani, (Eds.), Marcel Dekker, New York, 1980, pp. 321-347.

[10] B. Ren, *Comput. Chem.* **2002**, *26*, 357.

[11] B. Ren, *Comput. Chem.* **2002**, *26*, 223.

[12] B. Ren, *J. Mol. Struct. (THEOCHEM)*, **2002**, *586*, 137-148.

[13] B. Ren, *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 858-868.

[14] D. J. G. Marino, P. J. Peruzzo, E. A. Castro, A. A. Toropov, *Internet Electron. J. Mol. Des***. 2002**, *1*, 108-133, http://www.biochempress.com.

[15] A. A. Toropov and A. P. Toropova, *Russ. J. Coord. Chem.* **1998**, *24*, 81-85.

[16] A. A. Toropov and A. P. Toropova, *J. Mol. Struct. THEOCHEM* **2002**, 581, 11-15.

[17] A. A. Toropov, O. M. Nabiev, P. R. Duchowicz, E. A.Castro, F. Torrens, *J. Theor. Comp. Chem.* **2003**, *2(2)*, 139-146.

[18] D. J. G. Marino, P. J. Peruzzo, E. A. Castro, A. A. Toropov, *Internet Elect. J. Molec. Design* **2003**, *2*, 334-347.

[19] A. A. Toropov, P. R. Duchowicz, E. A. Castro, *Int. J. Mol. Sci.* **2003**, *4*, 272-283.

[20] P. R. Duchowicz, E. A. Castro, A. A. Toropov, *J. Arg. Chem. Soc.* **2002**, *90(1-3)*, 91-107.

[21] P. R. Duchowicz, E. A. Castro, A. A. Toropov, *Comp. Chem.* **2002**, *26(4)*, 327-332.

[22] P. J. Peruzzo, D. J. G. Marino, E. A. Castro, A. A. Toropov, *J. Mol. Struct. THEOCHEM* **2001**, *572*, 53-60.

[23] S. Wolfram, *The MATHEMATICA® Book*, Fourth Edition, Wolfram Media, Cambridge University Press, Cambridge, **1999**.

[24] H. L. Morgan, *J. Chem. Doc.* **1965**, *5*, 107.